Burstiness and Correlation Measures in Teletraffic

Master's Theses

by

György Miklós

Student of Informatics
Technical University of Budapest
Email: `MIKLOS@ttt-atm.ttt.bme.hu`

Supervisor:
Dr. Sándor Molnár
Department of Telecommunications and Telematics

Version 1.01, June 1997

Abstract

In this theses various measures of teletraffic burstiness are analyzed. The analysis is based on 15 traffic sources ranging from renewal type traffic to MMPP (Markov Modulated Poisson Process), overflow, aggregated, voice and video traffic models. Many commonly used burstiness measures are considered such as the squared coefficient of variation, peak to mean ratio, index of dispersion for intervals and counts and variance-time plot. The correlation structure of the traffic sources is also considered.

In the last part of the theses, the generalized peakedness of teletraffic processes is discussed and an application is presented for the estimation of delay in a queue with deterministic service.

Technical University of Budapest
Faculty of Electrical Engineering and Informatics
Department of Telecommunications and Telematics
High Speed Networks Laboratory

Contents

1 Introduction **5**

I Traffic sources **6**

2 Some comments on the traffic sources **6**

3 Renewal processes **7**
3.1 Smooth processes 8
3.1.1 Erlang inter-arrival time distribution 8
3.1.2 Uniform inter-arrival time distribution 8
3.2 Poisson process 8
3.3 Bursty processes 8

4 Correlated processes **9**
4.1 Linear processes 9
4.2 MMPP process 10

5 Overflow and aggregation **10**
5.1 Overflow traffic 10
5.2 Aggregated traffic 10

6 Traffic models of applications **11**
6.1 Voice traffic model 11
6.1.1 Introduction 11
6.1.2 Approximation of the length of intervals 11
6.1.3 Modeling the number of customer calls 11
6.1.4 Modeling the number of customers producing cells 12
6.1.5 Parameters of voice traffic 13
6.2 Video traffic model 13
6.2.1 Introduction 14
6.2.2 Time concepts 14
6.2.3 AR(1) model 14
6.2.4 IPP model 15
6.2.5 The video model 15
6.2.6 Setting the value of base unit 16

7 Summary of traffic sources **17**

II Traffic characteristics **18**

8 Characteristics based on inter-arrival times without sequential information 18
8.1 Squared coefficient of variation . . . 18
8.2 Peak to mean ratio . . . 18
8.3 Probability density function of inter-arrival times . . . 18
8.4 Moments of the inter-arrival times . . . 19

9 Correlation characteristics 19
9.1 Correlation of inter-arrival times . . . 19
9.2 Correlation of counts of arrivals . . . 19
9.3 Autocorrelation of counts of successive intervals . . . 19

10 Index of Dispersion measures 20
10.1 Index of dispersion for intervals . . . 20
10.2 Index of Dispersion for Counts . . . 20
10.3 Variance-time plot . . . 21

11 Connection between the correlation and dispersion index measures 21

12 Concluding remarks on traffic sources and characterization 23

III Traffic analysis 24

13 How to interpret the results 24

14 Discussion 25
14.1 Characteristics based on the inter-arrival times without sequential information . . . 25
14.2 Correlation characteristics . . . 26
14.3 Index of Dispersion measures . . . 28

15 Concluding remarks on burstiness characterization 30

IV Generalized Peakedness of Teletraffic Processes and Its Application in ATM Networks 32

16 Generalized peakedness 32

17 Evaluating generalized peakedness of an arrival process 32
17.1 Notations . . . 33
17.2 Evaluation of generalized peakedness . . . 34
17.3 Exponentially distributed holding times . . . 34
17.4 Asymptotic properties of generalized peakedness . . . 34
17.5 Generalized peakedness of sample sources . . . 35

18 Application of generalized peakedness **35**
18.1 Approximations of loss systems 35
18.2 Application to delay systems 35
18.3 Application of peakedness-based delay approximation in ATM: the $C_2/D/1$ queue 36
18.4 Simulation of $C_2/D/1$ queue 40
18.5 Comparison of generalized peakedness based waiting time approximation with simulation results 41

19 Concluding remarks on generalized peakedness **42**

A Characteristics of traffic models **45**

B Generalized peakedness of traffic sources **61**

C Summary of programs **70**

1 Introduction

In this theses the various burstiness measures used to characterize telecommunications traffic are examined. Burstiness is a very important aspect of traffic in ATM. It influences, among other things, cell loss ratio, buffer sizing, gain in statistical multiplexing.

I have examined the burstiness characteristics of 15 traffic sources. These sources range from simple renewal models to MMPP (Markov Modulated Poisson Process), overflow traffic, aggregated traffic, voice and video traffic models. The traffic sources were implemented in MATLAB. They all have mean intensity of one, thus they only differ in the shape of traffic, not the amount of traffic they produce.

Then I apply the various burstiness measures to these sources. I consider squared coefficient of variation, peak to mean ratio, histogram of inter-arrival times, index of dispersion for intervals and counts. I also consider the correlation structure of the sources.

The generalized peakedness of processes is defined. It is a very useful characterization of the variability of a teletraffic process since it takes into account the service characteristics and not just the process itself. An application is presented for the approximation of queueing delays in a $C_2/D/1$ queue.

This theses is organized as follows. In Part I the traffic models are introduced (with a summary in Section 7) In Part II the burstiness characteristics are presented (with a summary in Section 12). In Part III the results of applying the burstiness characteristics to the sources are discussed (with a summary in Section 15). Generalized peakedness is defined, discussed and an application is shown in Part IV. This part is summarized in Section 19. The values and graphs of the various characteristics applied to the traffic sources as well as graphs showing generalized peakedness values of the sources are shown at the end of the theses. A list of programs and references conclude the theses.

Acknowledgments

I began my studies in telecommunications traffic with a project laboratory at the Department of Telecommunications and Telematics, Technical University of Budapest [1]; my supervisor was Dr. Sándor Molnár. I continued in Lund, Sweden supported by TEMPUS at the Department of Communication Systems, Lund Institute of Technology [2] where my supervisor was Dr. Johan Karlsson. Later I took part in the Student Conference (TDK '96) [3] and at the Conference of Graduating Students [4] at the Technical University of Budapest, Faculty of Electrical Engineering and Informatics.

I thank the help of Dr. Sándor Molnár and Dr. Johan Karlsson. I also thank the help and interest of Attila Vidács.

The work presented in this theses is part of the traffic modeling research at the High Speed Networks Laboratory, Department of Telecommunications and Telematics, Technical University of Budapest.

Part I
Traffic sources

2 Some comments on the traffic sources

In later sections, I am going to present different types of traffic sources. These sources produce different types of traffic. There are 15 sources which I will compare in the analysis altogether.

I would like to make the following comments about these traffic source models.

- In this theses I deal with telecommunications traffic. Traffic can be thought of as a point process.

  A point process is made up of events on the infinite time axis. It is a stochastic process, the events are randomly distributed.

  A process can be considered from two different aspects [8]:

  Process of intervals between events. The process is described as a sequence of intervals between events. It is therefore a discrete description which consists of a sequence of positive numbers.

  Counting process. The process is described by the number of events in the interval $(0, t]$ as a function of t. It is therefore a continuous description which consists of a nondecreasing function. There is another description related to this: to divide the time axis into equal intervals, and to describe the process by the number of arrivals in each time unit. This is the counting process description as used in the practice.

- I will use the term process to refer to what is produced by a source. A process is made up of arrivals (of cells), which are events without any duration. In the case of traffic source models, I sometimes use the phrase 'production of cells'. The words 'process' and 'event' are terms used in time series analysis, the word 'arrival' and 'to produce cells' are more related to telecommunications.

- All of the sources have an average intensity of one, that is, the mean inter-arrival time is unity. This makes them comparable, the sources only differ in the shape of the traffic and not in the amount of traffic they produce.

- The models presented later in this study produce stationary processes. This is because there is no time dependence in the model from which they are produced. The only exception to this is the `voice` model, this will be commented on later in Section 6.1.

- The sources presented in this section use the inter-arrival (intervals between events) description. That is, their output is a sequence of numbers that corresponds to the times between consecutive arrivals. The `voice` and `video` models presented in the following sections use the other aspect first: they produce a sequence of numbers which correspond to the number of arrivals in consecutive fixed intervals. This description is then converted to inter-arrival description in such a way that arrivals in one interval are spread uniformly. This follows that all traffic sources use inter-arrival description in the end. This makes it possible to use the same analysis methods for all the sources.

- It is difficult for a human to visualize the sequence of inter-arrival times. That is why I usually plot the time-intensity graph to present a traffic. This is carried out as follows. A *time unit* is chosen, the time is quantized into periods of the length of this unit. For each period, the number of arrivals in that period is counted, and is divided by the unit length to get an intensity estimation for this interval. The choice of the base unit is, of course, difficult. This choice determines the time scale at which we look at the traffic.

- *The sources described below produce arrivals in 'unslotted' (continuous) time.* In the framework of ATM, however, cells can only arrive in slots which makes the arrival times discrete. The effects of slotted time axis are not considered here. I note that these effects become negligible if the link capacity is much larger than the rate at which the cells are produced (and so a slot is much shorter than the periods between arrivals.) Anyway, it should be kept in mind that the cell arrival rate is bounded from above if we have a slotted time axis.

- I have made simulations which realize all of these sources. The traffic source generators are implemented in MATLAB. The figures in this section come from these simulations. The results of the simulations are used later for the analyses of the traffic sources.

- The length of these simulations is 10000 inter-arrivals. (The average of these inter-arrival times is 1, as mentioned earlier, so that the mean intensity is fixed to be 1.)

3 Renewal processes

Renewal processes are characterized by independent identically distributed inter-arrival times. They are favored because of their simplicity. (Renewal processes are often used in approximations.)

Depending on the distribution of inter-arrival times, a process is said to be *smooth* or *bursty* if the distribution has a small or large variance, respectively. Quantitatively this can be described by the squared coefficient of variation of the inter-arrival time:

$$C^2(X) = \frac{Var(X)}{E^2(X)}$$

The process is said to be smooth if $C^2(X) < 1$ and bursty if $C^2(X) > 1$.

In the subsections that follow I describe the renewal sources that I use in my analysis. I first present the smooth processes, then the Poisson process ($C^2(X) = 1$), then the bursty processes.

3.1 Smooth processes

I consider two types of smooth renewal processes here: processes with Erlang and uniform inter-arrival distribution functions.

3.1.1 Erlang inter-arrival time distribution

Erlang distributions have long been used in telecommunications network design. A random variable from an Erlang-n distribution is the sum of n independent random variables of exponential distribution of parameter $n\lambda$. It follows easily that the intensity of a process with Erlang-n inter-arrival distribution is λ and its squared coefficient of variation is $1/n$. I realized an Erlang-2 and an Erlang-10 process, for which the squared coefficient of variation is 1/2 and 1/10, respectively. (The processes are later referred to as `erlang2` and `erlang10`.)

3.1.2 Uniform inter-arrival time distribution

I have also realized a renewal traffic source with uniform inter-arrival distribution on $[0, 2]$. In this case, the distribution has no tail: this type of source is never silent for more than 2 time units. The squared coefficient of variation evaluates to 1/3 in this case. (The process is referred to as `unifint`.)

3.2 Poisson process

The Poisson process is the simplest and the most basic of all processes. It is made up of arrivals where the inter-arrival times are exponentially distributed independent random numbers. The process can be characterized by the mean inter-arrival time, $1/\lambda$, or by the mean intensity, λ.

3.3 Bursty processes

In the case of the bursty renewal processes, the squared coefficient of variation is greater than one. I have realized two such processes whose inter-arrival distributions are hyper-exponential. In my case, the distribution is derived from two exponential distributions: it is exponential with parameter either λ_1 or λ_2, with probability p and $1-p$, respectively. By evaluating the first two moments of the hyper-exponential distribution, $m_1 = p/\lambda_1 + (1-p)/\lambda_2$, $m_2 = 2p/\lambda_1^2 + 2(1-p)/\lambda_2^2$, a formula for the squared coefficient of variation can be derived:

$$C^2(X) = 1 + 2\frac{p(1-p)(1/\lambda_1 - 1/\lambda_2)^2}{(p/\lambda_1 + (1-p)/\lambda_2)^2}$$

.

I have realized the following two renewal processes with hyper-geometrical inter-arrival distribution: The first has $p = 0.25$, $\lambda_1 = 0.2897$, $\lambda_2 = 5.475$, which gives 1 for the mean and 5 for the squared coefficient of variation; this process is referred to as `hyper5`. The second has $p = 0.1$, $\lambda_1 = 0.1358$, $\lambda_2 = 3.4126$, which gives 1 for the mean and 10 for the squared coefficient of variation; this process is referred to as `hyper10`.

We may think of the hyper-exponential process as an interrupted Poisson process: assuming $\lambda_2 > \lambda_1$ and $p < 1/2$, the second exponential distribution with higher intensity is more likely to be chosen. The infrequent choice of the first exponential distribution interrupts the sequence of inter-arrivals from the second distribution which forms a Poisson process.

4 Correlated processes

In contrast to the renewal processes of the previous section, correlated processes are made up of correlated inter-arrivals. Two types of correlated processes are presented here: linear processes and an MMPP process. Some of the processes described in later sections are also correlated.

4.1 Linear processes

Linear processes are useful when analyzing the correlation properties of a process [8].

When measuring a property of processes, it is desirable to apply the measurement to a process where the property is preset so that the accuracy and applicability of the measurement can be tested.

It is not possible to simply prescribe the correlation properties of a process like we can prescribe the intensity of a Poisson process. There is a way, however, to do this indirectly: by defining linear processes for which the correlation coefficients can be computed.

A linear sequence is defined as follows. A renewal sequence (defined by inter-arrival times ε_i) is convolved with a sequence of linear coefficients:

$$X_i = \sum_{s=-\infty}^{\infty} w_s \varepsilon_{i-s}$$

where $\sum w_s$ and $\sum w_s^2$ are finite. The correlation coefficients $(corr(X_i, X_{i+j}))$ are then given by

$$\rho_j = \frac{\sum_{s=-\infty}^{\infty} w_s w_{s+j}}{\sum_{s=-\infty}^{\infty} w_s^2}.$$

In practical situations, only a finite number of w_s are nonzero.

The sources (referred to as `linear1` and `linear2`) are linear processes with the coefficients listed in Table 1, which also shows the correlation coefficients. (All the other coefficients are zero.) We have positive lag 1 correlation for `linear1` and negative lag 1 correlation `linear2`.

	w_{-1}	w_0	w_1	ρ_0	ρ_1	ρ_2
`linear1`	1	1	1	1	2/3	1/3
`linear2`	1	-1	1	1	$-2/3$	1/3

Table 1: Coefficients of linear processes.

The generator renewal process has uniform distribution on $[2/9, 4/9]$ in the case of `linear1` and $[8/9, 10/9]$ in the case of `linear2` so that the resulting linear process will have the mean inter-arrival time of 1 and no negative inter-arrivals will result from the definition.

4.2 MMPP process

Markovian models are often used to describe a source with varying rates. A Markov Modulated Poisson Process (MMPP) is a Poisson process whose rate is determined by the state of a Markov process [19]. (So an MMPP can be described by the following: a finite number of states, a matrix of transition probabilities from one state to another, a mean sojourn time describing the exponential sojourn time distribution for each state, and a rate for each state.)

I have realized a Markov Modulated Poisson Process (referred to as `MMPP`) with two states: A and B. State A has a mean sojourn time of 10 and cells are produced at a rate of 5; state B has a mean sojourn time of 80 and cells are produced at a rate of 1/2. (This results in an average intensity of 1.)

5 Overflow and aggregation

5.1 Overflow traffic

The next two traffic sources are overflow processes from a queue. In the first case (referred to as `overflow1`) the queue is M/M/1/5: the incoming traffic is Poisson with intensity 1, the service time is exponential with intensity 1.1, the queue is limited to 5. In the second case (referred to as `overflow2`) the queue is M/M/1/8: the incoming traffic is Poisson with intensity 1, the service time is exponential with intensity 1.1, the queue is limited to 8.

When the queue is full, arrivals overflow. These arrivals make up the overflow process. (At the end of the simulation, the process is normalized to have intensity 1.)

This is a very realistic traffic. It is of bursty nature because when the queue is full, the next arrival is more likely to see it full.

5.2 Aggregated traffic

Superposition traffic plays a big role in ATM networks, that's why I investigate aggregate traffic. I multiplex 10 and 100 independent `hyper5` processes to get the next two traffic sources. (Referred to as `sup10hyp5` and `sup100hyp5`, respectively.)

I note here that it is important when super-positioning a lot of traffic sources that the individual sources are simulated from an arbitrary (random) time rather than an arbitrary arrival. Starting from an arbitrary arrival introduces non-stationarity to the beginning of the simulation [8]. In the case of an individual source this may be negligible, but as soon as hundreds of sources are multiplexed this kind of non-stationarity causes a transition period in the superposition which is not at all negligible.

6 Traffic models of applications

6.1 Voice traffic model

In this subsection I present a voice traffic model. An example of the time-intensity plot can be found in Appendix A under the page `voice`, on the top left. (See [3] for more details on the model.)

6.1.1 Introduction

Telephone voice traffic is made up of a superposition of many digitalized speech sources. I model a channel which is used by a great number of customers, each customer making a call in only a small percentage of time. When a customer makes a call, it is supposed to be a conversation. This is characterized by talking intervals and listening intervals. The line is not occupied under listening intervals. But even when the customer is talking, there are short pauses between sentences, words.

To sum it up, we model one source corresponding to a customer with traffic interrupted by three kinds of silences: inter-call, inter-talk, inter-word. The talk-spurts between these silences are modeled by constant cell rate. A standard voice bitrate is 64 kbits/sec, this can be translated to constant cell/sec terms depending on the cell size. (We might not want to wait until the whole cell is filled because of delay considerations. Filling up the whole ATM cell causes approximately 6 ms of delay.)

The traffic generated by one customer is very bursty; most of the time we have silence interrupted by occasional calls. But as soon as we model the channel, where we have lots of such sources multiplexed, the traffic will be averaged out to become much smoother. So the resulting traffic will be characterized by an average cell rate with some fluctuations added onto it.

Next I give a model for the superposition of many telephone voice traffic. I am going to use a discrete time approach, that is, I will have a *base unit* (defined in terms of seconds) and I will give a model which produces a sequence of numbers approximating the number of cells in a base unit for the superposition traffic.

6.1.2 Approximation of the length of intervals

In this model I use geometric distributions to model the length of all types of intervals. This is in agreement with measurements [29] [30], as far as the length of inter-talk, inter-word silence, talk spurt intervals are concerned. Inter-call silence and call length are also modeled by geometric distribution; this is probably not very far from reality.

Using geometric distributions I can make advantage of the memoryless property, that is, the probability that an interval is over in a base unit is given by the parameter of the geometric distribution, and these probabilities are independent over the sequence of base units. This simplifies calculations a lot.

6.1.3 Modeling the number of customer calls

I use a two-level approach: first I build a model for the number of customer calls, $N(t)$, out of the total number of customers M. Having the number $N(t)$, I go on to approximate

$m(t)$, the number of customers actually creating cells out of the $N(t)$; this is less than $N(t)$ because some customers may be in the state of inter-talk or inter-word silence.

I use two parameters to characterize the behavior of a customer as regards making calls: T_{call} which is the average call length, and λ which is the traffic offered by one customer. Denoting the average silence between two calls by T_{sil}, we have

$$\lambda = \frac{T_{call}}{T_{call} + T_{sil}}$$

which gives $T_{sil} = \left(\frac{1}{\lambda} - 1\right) T_{call}$. Time is measured here in seconds; let us use ΔT to denote the length of a base unit, then $T_{call}/\Delta T$ and $T_{sil}/\Delta T$ are measured in base units. It follows that the parameter of the geometric distribution of the call length measured in base units is given by

$$q = \frac{1}{T_{call}/\Delta T}$$

and that of the silence length is given by

$$s = \frac{1}{T_{sil}/\Delta T}.$$

(The geometric distributions are on 1,2,...)

At the $(t-1)$th base unit we have $N(t-1)$ customers calling out of the M. Each of those $N(t-1)$ customers finish their call with a probability of q in one base unit. It follows that the number of customers finishing their call in one base unit has a binomial distribution of order $N(t-1)$ and parameter q. Similarly, the number of customers beginning a new call is binomially distributed of order $M - N(t-1)$ and parameter s. Therefore,

$$N(t) = N(t-1) + B_s^{M-N(t-1)} - B_q^{N(t-1)},$$

where B_p^n is a binomially distributed random number of order n, parameter p.

6.1.4 Modeling the number of customers producing cells

We can approximate the number of customers producing cells very similarly to the approximation of number of customers making a call. Out of the possible $N(t)$ customers who are making a call, some are not producing cells because they are in the state of inter-talk or inter-word silence.

The only significant difference here compared to the previous subsection is that now we have two distributions for silence instead of one. I select one out of the two in a random manner: with probability c_1 we are in an inter-word silence, with probability c_2 we are in an inter-talk silence, $c_1 + c_2 = 1$. The actual values of c_1 and c_2 may have a cultural dependence, they determine how frequently the speakers interchange between speaking and listening.

The formulas for this part are give below. $T_{inter-word}$ is the average inter-word silence time; $T_{inter-talk}$ is the average time devoted to listening to the partner, $T_{talk-spurt}$ is the average time of a talk-spurt. The parameters of the corresponding geometric distributions are:

$$r_1 = \frac{1}{T_{inter-word}/\Delta T}, \; r_2 = \frac{1}{T_{inter-talk}/\Delta T}, \; p = \frac{1}{T_{talk-spurt}/\Delta T}.$$

At base unit t, we have $N(t)$ customers making a call, $m(t)$ of them are producing cells. Each of the $N(t) - m(t)$ customers which are in silence, they are in inter-talk silence with probability c_2. So the total number of customers which are in inter-talk silence is binomially distributed: $n(t) = B_{c_2}^{N(t)-m(t)}$. Then

$$m(t+1) = B_{r_2}^{n(t)} + B_{r_1}^{N(t)-m(t)-n(t)} - B_{p}^{m(t)}$$

6.1.5 Parameters of voice traffic

Next I show an example of voice traffic generated by the model with typical parameters. The following table describes the parameters used in the simulation later on.

parameter	*value*	*description*
ΔT	0.05 sec	base unit
$T_{inter-word}$	0.124 sec	inter-word silence mean time
$T_{inter-talk}$	4 sec	inter-talk silence mean time
c_2	1/17	probability of inter-talk silence
$T_{talk-spurt}$	0.17 sec	talk-spurt mean time
T_{call}	10 sec	mean call length
λ	0.1	traffic offered by one customer
M	1000	total number of customers
`len`	2000	length of the simulation in base units
`bitrate`	64000 bps	bitrate of digitalized voice
`cellength`	384 bit	length of a cell

It is interesting to note that the source model starts from zero state, that is, there are originally no calls in the system. It then takes some time to reach the steady state. In the analysis, however, this initial transition is removed to get a stationary source.

6.2 Video traffic model

In this subsection I discuss video traffic modeling. I present an AR(1)+IPP model [1]. After a short introduction and discussion about AR(1) and IPP models, I present the mathematical model. I also discuss some aspects of setting the parameters of the model. An example of the time-intensity plot can be found in Appendix A under the page `video`, on the top left. (See [3] for more details on the model.)

[1] Note that this model is not suitable for modeling MPEG video traffic. MPEG is characterized by 3 different types of frames, each having different traffic properties, with periodicity in the sequence of frame types.

6.2.1 Introduction

Video traffic has the characteristic of having very different bit rates when the scene is changing and when there is no change. These scene changes occur randomly.

This model uses the superposition of two processes to model this phenomenon. The AR(1) (Auto Resgressive) process is supposed to characterize the traffic without scene changes. This process is able to capture the local correlation properties of the process. The IPP is supposed to characterize the scene changes. This process exhibits a sudden increase in intensity for a while, and no intensity between these periods.

The model described here is mainly based on [26].

6.2.2 Time concepts

Here are descriptions of the time units which are to be used.

slot In the end, the model will have to produce ATM compatible results. There we have *slots* as time units, each slot is either empty or contains a *cell*. The length of a slot is Δt, whose value is not specified.

base unit The model will be in discrete time. That is, the model will produce a sequence of numbers, each meaning the number of cells in a *base unit*. This base unit is the smallest time period the model deals with; directly it says nothing about how the cells are distributed within a base unit. However, the best thing is to distribute the cells within a base unit uniformly to get a smooth traffic. The base unit is specified in terms of a slot; it always contains an integer number of slots.

frame A video sequence is made up of *frames*, it is typical to have 30 frames in a second. In this model we suppose that we are dealing with a coding scheme which uses some kind of forward prediction. Therefore we have two types of frames: *ordinary* frames and *scene change* frames. For ordinary frames, it is enough to send some kind of difference between the current and previous picture, and so the amount of information in an ordinary frame is low. For the scene change frames it is required to send the whole frame, therefore the amount of information contained in such frames is high. The frame length is specified in terms of the base unit; it always contains an integer number of them. It follows that one frame always contains the same number of slots regardless of its type, only the number of cells will be different.

6.2.3 AR(1) model

This is a discrete time recursive model. The number of arriving cells in a base unit is given by

$$x'(t) = y(t) + \rho$$

$$y(t) = ay(t-1) + e(t)$$

where

$x'(t)$ is the number of arriving cells in base unit t.

$y(t)$ is a random variable of the AR(1) process; its mean is 0.

ρ is the mean arrival rate of the AR(1) process.

a is a coefficient specifying the correlation of lag 1, $|a| < 1$ is required for the process to be stable.

$e(t)$ is a sequence of independent normally distributed random variables having mean 0, variance σ^2.

6.2.4 IPP model

In the IPP [2] model we have two states: ON and OFF. In the ON state, cells are produced with intensity λ, in the OFF state, no cells are produced.

In the video model, the length of both of these states are integer multiples of the frame length. The length of an OFF state in terms of frames is geometrically distributed on $0, 1, 2, \ldots$ with parameter p, the average length is therefore $1/p - 1$. The length of an ON state — that is, the length of a scene change — is always 1, except of course when the length of an OFF state between two ON states is 0.

The intensity of the IPP process is denoted by $w(t)$ here, its value is λ in ON states and 0 in OFF states. Since the frame length is a multiple of base unit, all ON and OFF states begin and end at base unit boundaries.

6.2.5 The video model

The video model is the superposition of the AR(1) and IPP model:

$$x(t) = y(t) + \rho + w(t)$$

$$y(t) = ay(t-1) + e(t)$$

As an output, this model produces the sequence of numbers $x(t)$ corresponding to the number of cells in base unit t. Then $x(t)$ must be rounded to the nearest integer, and this number of cells must be distributed evenly over the base unit.

To implement the model, the parameters a, σ, ρ, p and λ are needed. During the calculations, it is necessary to always have the value of $y(t-1)$, to know whether we are in ON or OFF state, and to know how many base units there are left in the current state.

Next, I give a set of parameters which can be used more easily instead of the ones above, I also give the transformation from the new parameters to the original ones.

The new parameters are:

f is the frame length in terms of base units.

m_{ord} is the mean number of cells in an ordinary frame.

σ_{ord} is the standard deviation of the number of cells in an ordinary frame.

[2] It is not the best to call this process IPP since that refers to a continuous-time model. It is also not necessary to use exponentially or geometrically distributed time periods for the stay in ON and OFF states. But since in [26] they use the term IPP, I did not modify it.

a is the correlation in the number of cells between two consecutive base units within an ON or OFF state. (When switching between ON and OFF states, this correlation does not apply.)

m_{sc} is the mean number of cells in a scene change frame.

l is the mean number of ordinary frames between two scene change frames.

Then the original parameters can then be expressed [3] as

$$\begin{aligned}
a &\quad \text{remains the same} \\
\rho &= \frac{m_{ord}}{f} \\
\sigma &= \sqrt{\frac{1-a^2}{f}}\sigma_{ord} \\
\lambda &= \frac{m_{sc}}{f} - \rho \\
p &= \frac{1}{l+1}
\end{aligned}$$

6.2.6 Setting the value of base unit

To determine which value to use for base unit, we have to take into account the following aspects.

Truncation effects We have to remember that we have a mathematical model, which produces a sequence of numbers with the prescribed statistical properties without imposing any kind of threshold on these numbers. Therefore it might well happen that for the number of cells in a base unit, we get a negative number or a number greater than the number of slots in a base unit. This is more likely to happen when we have few slots in a base unit, when we have high standard deviation for the process, or when the average number of cells in a base unit is set near to 0 or the number of slots. We also have to remember that the sequence $x(t)$ is made up of real numbers, which have to be rounded first to the nearest integer. This also produces an error in the average and standard deviation, especially if we have just a small number (etc. $1 \ldots 4$) of slots in a base unit.

To sum up all of these, we can decrease truncation effects by increasing the number of slots in a base unit.

Obscure correlation concept The value a was defined as the lag 1 correlation of the AR(1) process. It follows that this is for the correlation over 1 base unit, and *not* over one slot and *not* over one frame. To get a clear concept for a, we should set either the number of slots in a base unit to 1, or the number of base units in a frame to 1. However,

[3] In [26], they give a formula for a and σ^2 in terms of the mean of the AR(1) process (X_1) ρ_{X_1}, and $R_{X_1}(0)$, $R_{X_1}(1)$ where $R_{X_1}(i) = E[X_1(t)X_1(t+i)]$. Unfortunately both of these formulas are mistyped, the correct ones are: $a = \frac{R_{X_1}(1)-\rho^2_{X_1}}{R_{X_1}(0)-\rho^2_{X_1}}$ and $\sigma^2 = \frac{(R_{X_1}(0)+R_{X_1}(1)-2\rho^2_{X_1})(R_{X_1}(0)-R_{X_1}(1))}{R_{X_1}(0)-\rho^2_{X_1}}$.

the first thing would give high truncation effect, the second would result in too simple uniform traffic within one frame.

Keeping these ideas in mind, we have a tradeoff between truncation and obscure correlation. I think the best thing is to set the base unit to somewhere between a slot and a frame, so that a base unit is at least 4 slots.

However, if we model a simple codec which produces uniform traffic within one frame, then life is simpler. We can have the base unit equal to the frame length, and we have a clear correlation concept. In the simulations the parameters are set in this way.

7 Summary of traffic sources

All sources have unit intensity.

`erlang2` Renewal process with Erlang-2 inter-arrival distribution.

`erlang10` Renewal process with Erlang-10 inter-arrival distribution.

`unifint` Renewal process with uniform inter-arrival distribution on $[0, 2]$.

`poisson` Poisson process.

`hyper5` Renewal process with hyper-geometrical distribution, $p = 0.25$, $\lambda_1 = 0.2897$, $\lambda_2 = 5.475$, which gives 5 for the squared coefficient of variation.

`hyper10` Renewal process with hyper-geometrical distribution, $p = 0.1$, $\lambda_1 = 0.1358$, $\lambda_2 = 3.4126$, which gives 10 for the squared coefficient of variation.

`linear1` Linear process; correlation for lag 1 is 2/3, for lag 2 is 1/3. The generator renewal process has uniform distribution on $[2/9, 4/9]$.

`linear2` Linear process; correlation for lag 1 is -2/3, for lag 2 is 1/3. The generator renewal process has uniform distribution on $[8/9, 10/9]$.

`MMPP` Markov Modulated Poisson Process; State A mean sojourn time is 10, mean cell rate is 5; State B mean sojourn time is 80, mean cell rate is 0.5.

`overflow1` Overflow traffic from an M/M/1/5 type queue. Incoming traffic intensity is 1, service time intensity is 1.1.

`overflow2` Overflow traffic from an M/M/1/8 type queue. Incoming traffic intensity is 1, service time intensity is 1.1.

`sup10hyp5` The superposition of 10 `hyper5` sources, normalized to have intensity 1.

`sup100hyp5` The superposition of 100 `hyper5` sources, normalized to have intensity 1.

`voice` Voice traffic model

`video` Video traffic model

Part II

Traffic characteristics

8 Characteristics based on inter-arrival times without sequential information

8.1 Squared coefficient of variation

The squared coefficient of variation is defined as the ratio of the variance of the inter-arrival times to the square of expected inter-arrival time:

$$C^2(X) = \frac{Var(X)}{E^2(X)}$$

where X denotes the inter-arrival time as a random variable.

This is a simple burstiness characteristic which is based on the set of inter-arrival times, i.e. it does not take into account the order of arrivals. Although neglecting the order of arrivals makes the computation easy, it also limits the use of this characteristic: this measure only captures the variability of inter-arrivals, but the burstiness phenomenon is more complex: the order of arrivals is a very basic element of it.

8.2 Peak to mean ratio

The peak intensity to mean intensity is a very often used burstiness measure. It is defined as the peak intensity divided by the mean intensity. The meaning of the mean intensity is straightforward (total number of arrivals divided by the length of the time period considered), the meaning of the peak intensity on the other hand is quite hard to define. What is a peak? In this theses, the simplest notion is used: the peak intensity is determined by the two arrivals closest to each other in time.

I note that in slotted time (as in ATM, for example), the closest packet arrival from a source without traffic shaping is very likely to be one slot apart, so the peak intensity may not convey too much information. Other definitions of peak intensity are also possible, we could consider longer time periods. But the longer time units we take, the closer we get to the mean intensity—that's why the peak intensity is hard to define and use.

8.3 Probability density function of inter-arrival times

The distribution of inter-arrival times is another characteristic which is important in describing burstiness. The estimation of the pdf takes the inter-arrival times as a set of numbers without taking into account the order of them. The probability density function of the inter-arrivals can be approximated by its histogram.

When examining burstiness, it is important to look at the tail of the distribution. A very long (slowly decaying) tail implies occasional long silences, which is an indication of burstiness.

8.4 Moments of the inter-arrival times

When characterizing burstiness, we try to have simple descriptions. Instead of taking the whole probability density function of the inter-arrival times, we want to have just some numbers as the description of it. To this end, we can take the first some moments of distribution. Since the "length" of the tail plays an important role, it is not always enough to take the first moment (m_1, the mean inter-arrival time) and the second moment (m_2, which is related to the squared coefficient of variation: $C^2(X) = \frac{m_2 - m_1^2}{m_1^2}$), the third and the fourth moments (m_3 and m_4) can be important, too.

9 Correlation characteristics

9.1 Correlation of inter-arrival times

In this section I consider the characterization of the sources using correlograms. These correlograms show the correlation structure of the inter-arrival time sequence $X_1, X_2, \ldots$ The correlation of the sequence for lag j is

$$\rho_k = \frac{E\left[\left(X_i - E(X)\right)\left(X_{i+k} - E(X)\right)\right]}{Var(X)}.$$

The choice of X_i is arbitrary because of the stationarity. By $E(X)$ and $Var(X)$ I denote the expected value and variance of the inter-arrival time distribution which is common for all inter-arrivals also due to the stationarity.

Since ρ_k is normalized by the variance, its value is between -1 and 1. It expresses how correlated two inter-arrivals (separated by $k-1$ others) are. A value close to 1 shows that they are big or small at the same time, a value close to 0 shows that they are independent, and a value close to -1 shows that the high value of one implies the low value of the other. 'Big' and 'small' are taken relative to the mean value.

9.2 Correlation of counts of arrivals

The correlation can be evaluated for the counting process, too. In this case, we first discretize time, i.e. divide it into intervals of equal length. For this, a time unit, τ must be set. After setting the time unit, the process is described by the sequence of arrival numbers in subsequent intervals. The correlation of this sequence for lag i, $\rho_i(\tau)$ characterizes the correlation structure of the counts of arrivals.

9.3 Autocorrelation of counts of successive intervals

There is another way to describe the correlation structure of the counting process. In this case, we do not divide the time axis into intervals of equal length. The autocorrelation function ($\rho(t)$) is a function of time and is defined as the correlation between the counts of events in two successive intervals of length t.

10 Index of Dispersion measures

10.1 Index of dispersion for intervals

The index of dispersion for intervals (IDI) characterization is an extension to both the squared coefficient of variation and correlation. The main idea behind it is that in bursty traffic, long and short inter-arrivals are grouped together. Therefore the variance of k inter-arrivals taken together is greater than k times the variance of the original inter-arrival series.

The index of dispersion for intervals [8] is defined as a sequence of numbers J_k as

$$J_k = \frac{Var(X_{i+1} + \ldots + X_{i+k})}{kE^2(X)}$$

with $k = 1, 2, \ldots$ and E(X) is the common mean. The choice of i is arbitrary in this definition due to the stationarity.

We have in the numerator the variance of the sum of k consecutive inter-arrivals. It is normalized by $kE^2(X)$. Notice that for a renewal process, $Var(X_{i+1} + \ldots + X_{i+k}) = Var(X_{i+1}) + \ldots + Var(X_{i+k}) = kVar(X)$ where $Var(X)$ is the common variance. This follows that $J_k = \frac{Var(X)}{E^2(X)} = C^2(X)$, the squared coefficient of variation for all k. Renewal processes thus have a constant IDI sequence, the Poisson process as a special case has a constant 1 IDI sequence. The definition implies that $J_1 = C^2(X)$, the squared coefficient of variation for all sources.

Note that the computation of IDI requires stationarity. In the case of non-stationary sequences the IDI can still be computed, but the result will depend on which formula is used (another formula is shown in Section 11.

10.2 Index of Dispersion for Counts

Similarly to the IDI of the inter-arrival times process, we can define an index of dispersion measure for the counting process called Index of Dispersion for Counts (IDC) [8]. It is defined as

$$I_t = \frac{V(t)}{E(t)} = \frac{V(t)}{tm}$$

where $V(t)$ and $E(t)$ are the variance and expected number of the arrivals in an interval of length t, and $E(t) = tm$, where m is the mean intensity of arrivals. The IDC curve is defined so that it reflects $V(t)$ normalized to give identically one for the Poisson process.

It is interesting to examine the behavior of the IDC at 0. Choosing t small enough so that the probability of more than one arrivals in t is negligible, we have in a period of t one arrival with probability mt and zero arrivals with probability $1-mt$ giving $mt(1-mt)$ for the variance of arrivals and $1 - mt$ for the IDC value. It follows that the IDC curve tends to zero:

$$\lim_{t \to 0} I_t = 1.$$

It is known [8] that for a renewal process, the IDC asymptotically tends to $C^2(X)$ (the squared coefficient of variation) and that

$$\lim_{t \to \infty} I_t = \lim_{k \to \infty} J_k.$$

10.3 Variance-time plot

The variance-time plot shows the variability of the counting process on different time scales. Suppose that we quantize time into intervals of length t. Only the number of arrivals in such intervals (N) is observed in order to compute the average intensity over that interval, N/t. The variance of this random variable is $V(t)/t^2$. After normalizing by the mean intensity, we get the variance-time plot [9, 10] as

$$\frac{V(t)}{mt^2}.$$

Note that this is just the IDC value divided by t — that's why it is considered here, among the index of dispersion measures.

The variance-time plot of the Poisson process is then $1/t$. This shows that the Poisson process behaves less and less variably as the time scale is increased.

I note here that the autocorrelation function for the counting process is also related to the variance-time plot and $V(t)$ [8, 11]:

$$\rho(t) = \frac{V(2t)}{2V(t)} - 1.$$

In fact, this is the way how I compute the autocorrelation function.

11 Connection between the correlation and dispersion index measures

The derivative of index of dispersion for intervals (IDI) is closely related to the correlation of the source. To show this, consider the numerator in the definition of IDI, which can be expressed using covariances [8]:

$$\begin{aligned} \mathrm{Var}(X_{i+1} + \ldots + X_{i+k}) &= \\ k\mathrm{Var}(X) + 2\sum_{l=1}^{k-1}\sum_{h=1}^{l} Cov(X_i, X_{i+h}) &= \\ k\mathrm{Var}(X) + 2\sum_{l=1}^{k-1}(k-l)C_l & \end{aligned}$$

where C_l is the lag l covariance. (This is how I estimate the IDI sequence from the simulations.)

By introducing correlations (ρ_i is the lag i correlation) instead of covariances, we get for the index of dispersion

$$\begin{aligned} J_k &= \frac{\mathrm{Var}(X)}{E^2(X)k}\left(k + 2\sum_{l=1}^{k-1}(k-l)\rho_l\right) \\ &= C^2(X)\left(1 + 2\sum_{l=1}^{k-1}\left(1 - \frac{l}{k}\right)\rho_l\right). \end{aligned}$$

The derivative (slope) of the IDI graph is then

$$J_{k+1} - J_k = C^2(X) \frac{\sum_{l=1}^{k} l\rho_l}{\frac{k(k+1)}{2}}.$$

In words, it is the squared coefficient of variation times a weighted average of the correlation coefficients from 1 to k, where the weights are just equal to the lag.

In a similar fashion, the derivative of index of dispersion for counts (IDC) is closely related to the correlation of the source, this time to the correlation of the counts of arrivals in successive equal intervals.

Let now denote τ the time unit, then the IDC value is defined as [8]

$$\begin{aligned} I_{k\tau} &= \frac{V(k\tau)}{m(\tau)k} \\ &= \frac{1}{m(\tau)k}\left(kV(\tau) + 2\sum_{l=1}^{k-1}(k-l)C_l(\tau)\right) \\ &= \frac{V(\tau)}{m(\tau)}\left(1 + 2\sum_{l=1}^{k-1}\left(1 - \frac{l}{k}\right)\rho_l(\tau)\right) \end{aligned}$$

where $C_l(\tau)$ is the lag l covariance for the sequence of arrivals in intervals of length τ, and $\rho_l(\tau)$ is the corresponding correlation; $m(\tau)$ is the mean number of arrivals in a period of τ.

The derivative (slope) of the IDC graph (with respect to k) is then

$$I_{(k+1)\tau} - I_{k\tau} = \frac{V(\tau)}{m(\tau)} \frac{\sum_{l=1}^{k} l\rho_l(\tau)}{\frac{k(k+1)}{2}}.$$

In words, it is the variance to mean ratio for the number of arrivals in a period of τ times a weighted average of the correlation coefficients from 1 to k, where the weights are just equal to the lag.

In a similar way, the autocorrelation function $\rho(t)$ of counts of successive intervals (see Section 9.3) can be expressed at integer multiples of τ. By definition, $\rho(k\tau)$ is the correlation between the counts N and N' of two successive intervals of length $k\tau$. Denoting the counts of intervals of length τ by $n_1, n_2, \ldots, n_k$ in the first interval of length $k\tau$ and by $n'_1, n'_2, \ldots, n'_k$ in the second interval of length $k\tau$, the covariance between N and N' reads as

$$\begin{aligned} C(\tau) &= E((N - E(N))(N' - E(N'))) \\ &= E(NN') - E(N)E(N') \\ &= E(NN') - E^2(N) \end{aligned}$$

since $E(N) = E(N')$ due to stationarity. Substituting $N = n_1 + n_2 + \ldots + n_k$, $N' = n'_1 + n'_2 + \ldots + n'_k$ and $E(N) = kE(n)$ (where $E(n) = E(n_1) = \ldots = E(n'_k)$) gives

$$\begin{aligned} C(\tau) &= E((n_1 + n_2 + \ldots + n_k)(n'_1 + n'_2 + \ldots + n'_k)) - k^2E^2(n) \\ &= C_1(\tau) + 2C_2(\tau) + \ldots + kC_k(\tau) + (k-1)C_{k+1}(\tau) + \ldots + C_{2k-1}(\tau) \end{aligned}$$

where $n_i n_{i+j} = C_j(\tau) - E^2(\tau)$ was used. To get to the correlation, we need $\text{Var}(N)$ which is (similarly as above)

$$\text{Var}(N) = k\text{Var}(n) + 2((k-1)C_1(\tau) + \ldots + C_{k-1})$$

This gives (after canceling by $\text{Var}(n)$)

$$\rho(k\tau) = \frac{\rho_1(\tau) + 2\rho_2(\tau) + \ldots + k\rho_k(\tau) + (k-1)\rho_{k+1}(\tau) + \ldots + \rho_{2k-1}(\tau)}{k + (k-1)\rho_1(\tau) + \ldots + \rho_{k-1}(\tau)}.$$

It shows that $\rho(k\tau)$ can be evaluated from the sequence of counts of arrivals, although the relationship is not at all straightforward.

12 Concluding remarks on traffic sources and characterization

In Part I and II the problem of characterizing the burstiness of teletraffic processes was introduced by presenting the wide range of characterization methods found in the literature and by presenting a variety of traffic sources to which these methods should be applied.

In Part I, the investigated traffic sources were presented. The selection of traffic models include: renewal models with smooth, Poisson and bursty inter-arrivals, linear and MMPP processes, overflow process, aggregated processes, voice and video models. All the traffic models (15 altogether) were implemented (in MATLAB) so that they all have a mean intensity of one. In this way, they only differ in the shape of the traffic they produce and not in the amount. These sources cover a wide range of traffic behavior.

In Part II various burstiness characterization methods are presented. The most simple ones are the squared coefficient of variation, peak to mean intensity ratio, probability density function of interarrival times. The correlation of the processes was also considered. Three different correlation measures were defined: the correlation in the sequence of interarrival times, the correlation in the sequence of counts in consecutive fixed periods, and the correlation between two consecutive intervals of length t. Finally, as an extension to both the squared coefficient of variation and correlation measures, the IDI, IDC and variance-time plots were defined and the connection between the indices of dispersion and correlation was shown. The last equation for the autocorrelation of the counts in successive intervals of length t was derived by the author.

In the next part the application of the characterization methods to the sample source models of Part I is discussed.

Part III
Traffic analysis

In this part I present the results of applying the above mentioned traffic characteristics to the sources presented in part I. The graphs are seen at the end of the theses (see Appendix A).

13 How to interpret the results

I would like to make the following comments on how the computations were made and on the interpretation of the results.

- Each traffic source was implemented in MATLAB. The output of the program realizing the models was a sequence of inter-arrival times. For the results below, this sequence was 10000 inter-arrivals long.
- Each computation was repeated ten times in order to establish confidence intervals. All results are shown as the average of the ten experiments with the 90% confidence intervals. On the graphs this is shown as two dotted curves above and below the solid curve showing the average, or as bars. The confidence intervals give an idea of the precision of the results. In the title of the graphs, "10 exp." refers to the number of experiments the average and the confidence intervals were computed from.
- For each traffic source, I give the estimated values of the first, second, third and fourth moments of inter-arrival times, the squared coefficient of variation of inter-arrival times, and the peak to mean ratio.
- Please note the scales on the axes of graphs. Although I tried to use as few scalings as possible, I could not use always the same because of the big differences of the graphs. Note that some graphs are logarithmic.
- I present an intensity-time plot for every source to give an idea how the traffic source behaves. As mentioned in section 2, this is done based on a time unit selection. A time unit is chosen, and the average intensity is calculated on successive intervals of that length.
- The estimation to the probability density function of the inter-arrival times is plot. The pdf is approximated by its histogram, using bins of equal length. The length of the bins can be seen on the staircase plot as the length of one 'stair'. The confidence intervals are plotted by two dots in order not to confuse the graphs.
- The estimation of correlation of the inter-arrival sequence is plot. The confidence intervals are shown by bars. Note that the correlation for lag zero is 1 by definition, this value is not always plot.

- The estimation of the counting process correlations is also plot. In this case, a time unit must be chosen. The value of this unit is seen in the title line of the graphs.

- The estimation of the IDI is plot. Although the IDI is defined for integer lags just as correlation mentioned above, the result is plotted as a continuous curve as it is usually done like this.

- The estimation of autocorrelation function of the counting process, the IDC curve and the variance-time plot are also shown together with confidence intervals as dotted lines. These curves are related to each other, in fact, each of them was computed from $V(t)$ as described in section 10.3.

- The variance-time functions also show the $1/t$ lines corresponding to the Poisson process. Values above this line imply greater variability, values below the line imply lower variability for a given time-scale (t).

- I used the following labels to identify the graphs: `intensity` for the intensity-time plot, `pdf` for the probability density function estimate of the inter-arrival length, `correlation` for the correlation in the sequence of inter-arrival times, `dcorrelation` for the correlation in the sequence of counts of arrivals in successive intervals of equal length, `IDI` and `IDC` for the index of dispersion for intervals and counts, `vartime` for the variance-time plot, and `autocorrelation` for the correlation in the number of arrivals in two successive intervals of length t.

14 Discussion

In this section I discuss the results of comparing the various burstiness measures.

14.1 Characteristics based on the inter-arrival times without sequential information

- The *squared coefficient of variation* is very easy to compute, and it gives us a measure of the variability of inter-arrival times. We can estimate it quite accurately. (We get back the prescribed values for the renewal processes within the confidence intervals.)

  However, these values should be handled with care: they can be misleading. Compare, for example, the estimated value for the `hyper5` process (5.03) to that of the source `video` (1.42): it contradicts the fact that the `video` source is very bursty. Or consider `voice`: it has a very low squared coefficient of variation (0.0181), about the same as for `linear2` (0.0123), but the two traffic look considerably different.

- We can look at burstiness as being caused by two factors [12]: the distribution of inter-arrival times and the correlation structure in the inter-arrival time sequence. The squared coefficient of variation ignores the correlation structure, and characterizes only the inter-arrival distribution. But it is not even enough for that. A very important factor in burstiness is the tail of this distribution. Even if they are infrequent, very long inter-arrivals imply burstiness (we keep now the average intensity constant) due to the long silence they imply. Two sources could have the same squared coefficients of variation, yet very different tails.

- The graphs on the inter-arrival probability density estimation show that this distribution can be estimated with relatively high accuracy.
- Notice that logarithmic plot is used on the y-axis. The slope of the pdf of `poisson` is -1 because it is exponential, while the slope of `erlang2` and `erlang10` tend to -2 and -10, respectively. We can observe that the Erlang distributions have "short tails" while the hyper-exponential distributions have "long tails" in general.
- Notice that the distributions of `linear1` and `linear2` are identical. However, the time-intensity plots are different due to the difference in the correlation structure.
- Observe that the pdf of inter-arrival distribution of `MMPP` and the hyper-exponential distributions look very similar, in fact, we cannot differentiate an MMPP and a hyper-exponential renewal process based on the inter-arrival distribution solely.
- Notice that for `sup10hyp5` and `sup100hyp5`, the inter-arrival distribution tend to that of the `poisson`. It does not mean, however, that those superposition traffic types are Poissonian, their correlation structure is different, as shown later.
- The inter-arrival distribution of `voice` and `video` is discrete, because the models originally use a time unit and distribute the generated number of arrivals in that time unit uniformly; this method only allows a discrete inter-arrival time distribution.

14.2 Correlation characteristics

- The correlation in the inter-arrival sequence shows the renewal property, i.e. that the inter-arrival times are independent, for the following sources: `erlang2`, `erlang10`, `unifint`, `poisson`, `hyper5`, `hyper10`, `overflow1`, `overflow2`. The first six sources were defined to be renewal, so it is not surprising that we get independent inter-arrivals.
- The overflow sources are also of renewal types, because the traffic arriving at the queue is renewal, and the queue just works as a "gate": it lets through some inter-arrivals, which will then not be part of the overflow, and it directs some inter-arrivals, which will form the overflow traffic. So the inter-arrivals in the overflow process come from the inter-arrivals of the process arriving at the queue. And since that was renewal, the overflow process is also renewal.
- For `linear1` and `linear2`, we get back the prescribed values accurately.
- The `MMPP` source shows positive correlation in the inter-arrival sequence for lags up to about 60. The is because the parameters were set so that in State A, 50 arrivals are produced, and in State B, 40 arrivals are produced on the average. It follows that a lot of consecutive short and a lot of consecutive long inter-arrivals are present in the sequence, which implies positive correlation.

- Observe the effect of super-positioning `hyper5` sources. In spite of the fact the `hyper5` source was renewal, the superposition process shows positive correlation in the inter-arrival sequence [11]. The more sources we superimpose, the longer the positive correlation takes. It shows that it is not feasible to model superposition traffic by a renewal model. To understand this phenomenon, we can think of the individual sources as IPPs: a Poisson process modulated by a two-state (ON and OFF states) Markov process. The superposition traffic intensity is determined by the number of IPPs in the ON state. This traffic intensity changes slowly (as compared to the inter-arrival times) which follows that many inter-arrivals of the same magnitude will be grouped.

- The `voice` source shows slowly decaying correlation in the inter-arrival sequence because of the long-term fluctuations in the intensity of the source.

- The `video` source is characterized by many consecutive long and many consecutive short inter-arrivals. Although the low intensity periods are much longer, the inter-arrival periods during low intensity are also longer. There are altogether about 40% of the inter-arrivals which are in the low intensity periods, which means that the number of short and long inter-arrivals are balanced. That is why it shows strong positive correlation for up to lag 200. The initial part (lag 1-500) is reminiscent of the correlation structure of a square-pulse like sequence, which shows a saw-wave like correlation structure.

- The correlation can also be estimated for the sequence of counts of arrivals. As will be seen below, this correlation measure cannot be computed from the correlation of the inter-arrival sequence alone.

- The sequence of counts shows negative correlation for the smooth renewal processes (`erlang2`, `erlang10`, `unifint`). The reason is that the inter-arrival times do not vary too much, and therefore if one arrival is "missing" in one time unit, it will be probably present in the next. That is, if the arrivals in one unit is smaller than the average, the next unit will probably have more arrivals, and vice versa.

- The Poisson process is not correlated with respect to the sequence of counts. As it is known from the theory of stochastic processes, the number of arrivals in consecutive intervals follow independent Poisson distributions.

- The bursty renewal processes show positive correlation in the sequence of counts. (It is best seen for `hyper10`. The sources `hyper5`, `overflow1`, `overflow2` are also in this category, but their coefficient of variation is not big enough.) It is explained by the bigger frequency of long inter-arrivals. A long inter-arrival implies many small consecutive values for the sequence of counts, thus giving positive correlation.

- The `linear1` and `linear2` processes show negative correlation in the sequence of counts (since they are smooth). The `linear2` source is more negatively correlated because its negative correlation in the sequence of inter-arrivals implies that a short inter-arrival will probably be followed by a long and vice versa, and then the same applies to the correlation of the sequence of counts.

- In the case of the `MMPP`, `sup10hyp5`, `sup100hyp5`, `voice` we can notice that we have to have positive correlation for big lags in order to have positive correlation in the sequence of counts of arrivals. For example, the MMPP inter-arrival sequence shows significant positive correlation for up to lag 50, which means more than 6 units (the time unit is 8). The correlation in the sequence of counts is positive for lag 1, and becomes zero at about lag 6.

- In the case of the `voice` source, we have significant positive correlation for lags up to about 200 in the sequence of counts. This translates to 1600 in the time axis. This value is the size of the long-term "waves" in the intensity-time plot.

- It is very interesting to note that the sequence of counts of arrivals for the `video` sequence only shows positive correlation up to about lag 7, which translates to 48 in the time axis. This is the length of a high activity period (approximately 50). In the correlation plot of the sequence of counts of arrivals we get positive values up to the lags which correspond to the large-scale changes in the intensity-time plot.

- As shown in Section 11, the autocorrelation between successive intervals of length t can be computed at $k\tau$ from the correlation of the sequence of counts of arrivals. However, this relationship is hard to visualize.

- In general, the autocorrelation curve shows nonzero values on the time scales on which the intensity-time graph has some kind of regularity. (See `video`, `sup10hyp5`, `sup100hyp5`, for example.)

- The linear sources have an interesting wavy autocorrelation function. In the case of `linear2`, the minimums at 0.5, 1.5, etc. can be noticed very clearly. The explanation is that the `linear2` has inter-arrivals very close to 1, and the negative correlation even intensifies the regularity of the sequence of arrivals. Therefore, in successive intervals of length 1.5, for example, the number of arrivals is 1,2,1,2, ... with very rare deviations from this sequence, thus giving negative correlation. In successive intervals of length 2, however, the sequence of arrivals is something like 2,2,2,... with rare increases and decreases, and this sequence does not show correlation. This periodicity in the autocorrelation curve suggests periodicity in the arrivals.

14.3 Index of Dispersion measures

- The IDI graphs are useful for traffic characterization because they show the variability of the traffic source on different scales. The IDI graphs can be regarded as an extension to both the squared coefficient of variation and the correlation measure of the sequence of inter-arrival times. The variability of the source on a given scale is shown by the value of the IDI graph, the correlation in the sequence of inter-arrival times is shown by the change in the IDI graph. (See Section 10.1 and 11.)

- For all the renewal processes (`erlang2`, `erlang10`, `unifint`, `poisson`, `hyper5`, `hyper10`, `overflow1`, `overflow2`) the IDI is constant, and its value reflects the squared coefficient of variation. The higher the constant value of the IDI, the more bursty the source is.

- For the linear processes, the IDI takes very low values implying low burstiness. The IDI for `linear1` increases showing positive correlation, the IDI for `linear2` decreases showing negative correlation.

- The IDI graphs for `MMPP`, `sup10hyp5`, `sup100hyp5` first increase then to stabilize at a value greater than one, thus showing burstiness. Notice that the superposition of `hyper5` sources to `sup10hyp5` and `sup100hyp5` does not change the final value of the IDI graphs significantly. The IDI for lag one, however, changes, as it is the squared coefficient of variation.

- The IDI graph of the `voice` source shows increase which means positive correlation. But the actual values are extremely inaccurate due to the inaccuracy in the correlation estimates. This shows the drawback of using IDI graphs: it takes a lot of computation, but the result may be extremely inaccurate.

- This IDI graph of `video` shows high burstiness both because of the quick increase and the stabilization at a high value.

- Comparing the IDI and IDC graphs we can notice that the graphs are really equal at infinity.

- For the renewal processes, the IDI is constant at the squared coefficient of variation. This follows that the IDC tends to the squared coefficient of variation, implying an increasing IDC when $C^2(X) > 0$ and a decreasing IDC when $C^2(X) < 0$. That is, the IDC is not constant for the renewal processes (as the IDI). The special case is the Poisson process: it has constant one IDI and IDC curve.

- As shown in Section 11, we can relate the derivative of the IDC to the correlation in the sequence of counts of arrivals. As can be seen on all the graphs, positive correlation implies increasing IDC, negative correlation implies decreasing IDC. Notice on the IDC of for example `erlang2`, `unifint`, `hyper5` that the IDC curve is very sensitive to even very little deviations from zero in the correlation.

- Notice that the results shown in Section 11 apply to the difference $I_{(k+1)\tau} - I_{k\tau}$, that is, only the values of IDC at multiples of τ are considered. (The IDC is a function of time as opposed to the IDI which is a function of lag.) In this case, tau is the time unit used to plot the correlation of the sequence of counts of arrivals.

- As can be seen in the cases of `MMPP`, `sup10hyp5`, `sup100hyp5`, `voice`, `video`, inaccuracies in the correlation estimates can lead to accumulated inaccuracies in the estimation of the IDC curve. It is important to estimate the IDC curves not only once, but several times in order to establish its precision — the IDC curve is prone to errors.

- The linear processes have an IDC with some waves. This can be understood as follows. The linear processes have very little varying inter-arrival times. (This is especially true for the `linear2` source because the negative correlation does not allow to accumulate the deviations.) Therefore, taking $t = 1$, for example, the number of arrivals is nearly always 1, so the variance is little. Taking $t = 1.5$, the number of arrivals is either 1 or two, thus increasing the variance to be much higher and consequently giving higher value for the IDC at 1.5 than at 1.

- Note that super-positioning `hyper5` sources, the resulting IDC will have the same value at infinity. More generally, super-positioning k independent identical sources, the IDC at t for one source is $I_t = \frac{V(t)}{mt}$, for the superposition traffic, $I_t^{(k)} = \frac{kV(t)}{kmt} = I_t$, that is, super-positioning does not change the IDC. It is only because the `hyper5` sources were not only super-positioned but normalized to give mean intensity one that the IDC curves are different.

- As described in Subsection 10.3, the variance-time plot is very much related to the IDC in that it is derived by dividing by t. Note, however, that by using logarithmic plots, the variance-time curves can be drawn over several time scales with much less relative error than the IDC. This method gives us a possibility to estimate the value of IDC at infinity by the slope of the variance-time plot at infinity much more precisely than from the IDC itself.

- The variance-time curves show that on larger scales, the traffic intensity fluctuations are averaged out and the traffic becomes smoother. This is not seen on the IDC curves.

- The $1/t$ line corresponding to the Poisson process is also shown. Values above this line imply more bursty sources, values under this line imply less bursty sources.

- All the traffic sources have a variance-time curve with its slope tending to -1, the same as for the Poisson process; there is only one exception, the `voice` source. This behavior can be attributed to the fact that the `voice` source shows fluctuations on the time scale of 2000 - 3000, and the variance-time is only plot to 1000.

15 Concluding remarks on burstiness characterization

In Part I, II, III the problem of characterizing the burstiness of teletraffic processes was investigated. A variety of possible characterization methods were applied to a wide selection of teletraffic models.

In Part III the results of applying the characterization methods to the sample sources is discussed. The plots and figures are presented at the end of the theses. The main results are summarized as follows.

- The squared coefficient of variation is easy to compute but it does not capture the burstiness propery well. Its main drawback is that it does not take into account the order of inter-arrivals.

- The peak-to-mean ratio cannot be used in general because it is not possible to give a general definition of a peak, moreover this ratio gives little information about the traffic.

- The probability density function of the inter-arrival times gives all the information about the inter-arrival time values as a set of numbers, but says nothing about the order of the inter-arrivals.

- The various correlation plots are all informative. The different correlation plots demonstrate different aspects of the correlation phenomenon.

- Correlation and burstiness are two different concepts. There are examples among the sample sources for relatively bursty and uncorrelated process (`hyper10`), not bursty and uncorrelated (`unifint`), bursty and correlated (`video`) and not bursty and correlated (`linear1`).

- Correlation is not even sufficient when only one type of traffic is considered because aggregating more traffic sources may increase correlation yet decrease burstiness.

- Index of dispersion measures are useful for burstiness characterization if both the absolute value and the slope of the graphs are taken into account.

- The drawback of using indices of dispersion is that they require a lot of computation and they may be very inaccurate.

In the next part, I discuss another traffic characterization method, namely the generalized peakedness. This concept takes into account both the arrival process and the service process of the system.

Part IV

Generalized Peakedness of Teletraffic Processes and Its Application in ATM Networks

In this part the concept of generalized peakedness is introduced (Section 16) its evaluation and some of its properties are discussed (Section 17) and an application to ATM network dimensioning is shown (Section 18). This part of the theses is concluded in Section 19.

16 Generalized peakedness

Peakedness was originally used in the approximate analysis of complex loss systems. It was meant to partially characterize a stream overflowing a finite group of servers. The peakedness of a stream overflown from a finite server group has been defined as the variance-to-mean ratio of the number of busy servers, in an infinite group of servers, to which this overflow stream is hypothetically offered [13]. It was implied in the definition that the holding times in both the original finite server group which is overflown and the infinite server group to which the traffic is offered are exponentially distributed.

One obvious way to extend the definition is to allow peakedness to be a characterization of any stationary arrival stream and not just overflow streams. Another generalization is to consider not just exponential holding times but any kind of distribution with a finite mean.

This leads to the definition of generalized peakedness [13] as the peakedness value (variance-to-mean ratio of the number of busy servers in an infinite group of servers) for a given stationary arrival stream as a function of the holding time distribution. The holding time distribution is usually given in terms of the complementary distribution function: $F(x) = P(\text{holding time} > x)$. *The generalized peakedness of a given teletraffic process is therefore a functional which maps a complementary holding time distribution F to a peakedness value $Z[F]$.*

The idea behind the definition is that the variability (i.e. burstiness) of the traffic in a system does not solely depend on the arrival process but is also determined by the holding times of the servers. The generalized peakedness value takes into account both the arrival stream and the process of the service system and it is intended to capture the interactions between the arrival and service processes.

17 Evaluating generalized peakedness of an arrival process

In this section some formulas for the evaluation of the generalized peakedness are summarized based on [13].

17.1 Notations

The following notations will be used.

S arrival stream (stationary).

$M(\cdot)$ mean-time curve of S: $M(t) = \mathrm{E}(N(t))$.

$V(\cdot)$ variance-time curve of S: $V(t) = \mathrm{Var}(N(t))$.

m intensity of S: m=M'(t), constant.

$U(\cdot)$ expectation function of S: $U(x)$ = expected number of arrivals in S in the time interval $(a, a+x]$ where a is an arbitrarily chosen arrival. ($U(x)$ is defined to be zero for $x < 0$.)

$\widetilde{U}(\cdot)$ Laplace-Stieltjes transform of U:

$$\widetilde{U}(s) = \int_{0-}^{\infty} e^{-sx} dU(x)$$

b batchiness parameter of S: the mean size of the batch that an arbitrary arrival finds itself in. If there are no batches (simultaneous arrivals) in S, then S is said to be orderly, and $b = 1$.

V/M is defined by $\lim_{t\to\infty} V(t)/M(t)$, which is the value of the IDC curve (and also the IDI) at infinity.

T is the holding time random variable associated with a server in the infinite server group to which S is hypothetically offered.

μ is the service rate of a server: $\mu = 1/\mathrm{E}(T)$.

$F(\cdot)$ complementary distribution function of T:

$$F(x) = \begin{cases} P(T > x) & \text{if} \quad x \geq 0 \\ 0 & \text{if} \quad x < 0. \end{cases}$$

$F_2(\cdot)$ autocorrelation of F:

$$\begin{aligned} F_2(x) &= F(x) * F(-x) \\ &= \int_{-\infty}^{\infty} F(u)F(u-x)du. \end{aligned}$$

$L(t)$ number of busy servers in the infinite group at time t.

$Z[\cdot]$ peakedness functional of S: $Z[F] = \mathrm{Var}(L(t))/\mathrm{E}(L(t))$, which is independent of t due to the stationarity of S.

$G(\cdot)$ normalized complementary distribution associated with F: $G(x) = F(x/\mu)$ so that it corresponds to unit mean holding time.

$z_G(\cdot)$ peakedness function for the 1-parameter family associated with G, that is, the distribution given by G is scaled to have mean μ: $z_G(\cdot) = Z[F]$, where $F(x) = G(\mu x)$. This gives the peakedness of the arrival stream as a function of the mean holding time. The shape of the distribution is fixed, only the mean varies.

17.2 Evaluation of generalized peakedness

Using the notation above, the generalized peakedness of an arrival stream can be expressed in terms of the expectation function and the autocorrelation of the complementary holding time distribution as [13]

$$Z[F] = 1 + 2\mu \int_{0-}^{\infty} F_2(x) dU(x) - m/\mu.$$

For the 1-parameter family of holding time distributions parametrized by the mean,

$$z_G(\mu) = 1 + 2 \int_{0-}^{\infty} G_2(\mu x) dU(x) - m/\mu.$$

It can be shown that for a Poisson process, $Z[F] = 1$ for all F. Generally speaking, $Z[F] > 1$ when S is bursty and $Z[F] < 1$ when S is smooth.

17.3 Exponentially distributed holding times

The special case when the server holding time is exponentially distributed is very important. The formulas in this case are much simpler. Furthermore, it can be shown [13] that the function $z_{\text{exp}}(\cdot)$ gives a complete second order characterization of the arrival stream, and so $Z[\cdot]$ can be expressed in terms of $z_{\text{exp}}(\cdot)$.

When T is exponentially distributed, $F_{\text{exp}}(x) = e^{-\mu x}, x \geq 0$, the generalized peakedness for exponential holding time with parameter μ is simply

$$z_{\text{exp}}(\mu) = 1 + \widetilde{U}(\mu) - m/\mu.$$

The expectation function U and its Laplace-Stieltjes transform can be evaluated for several important traffic models, and so the evaluation of generalized peakedness in the case of exponential holding times is not too difficult.

If the function $z_{\text{exp}}(\cdot)$ is already known, it can be used to determine the value of $Z[F]$. In [13] $Z[F]$ is expressed in terms of $z_{\text{exp}}(\cdot)$ in the case when F is of the form $F(x) = \int e^{-xy} a(y) dy$ where $a(y)$ is a generalized function. This method is elaborated and a formula is given for the case of Coxian holding time distributions in [15].

17.4 Asymptotic properties of generalized peakedness

If only the service rate μ of the holding time distribution is changed, the generalized peakedness for this 1-parameter set of distributions $z_G(\mu)$ has interesting asymptotic properties [13]:

$$\begin{aligned}\lim_{\mu\to 0+} z_G(\mu) &= 1 + G_2(0)(V/M - 1)\\ \lim_{\mu\to\infty} z_G(\mu) &= 1 + G_2(0)(b-1)\end{aligned}$$

If the arrival process S is orderly (that is, there are no simultaneous arrivals), then $b = 1$ and therefore $\lim_{\mu\to\infty} z_G(\mu) = 1$.

For deterministic holding times, $G_2(0) = 1$ and so $\lim_{\mu\to 0+} z_G(\mu) = V/M$. For exponential holding times, $G_2(0) = 1/2$ and so $\lim_{\mu\to 0+} z_G(\mu) = (V/M + 1)/2$.

17.5 Generalized peakedness of sample sources

The generalized peakedness values for certain holding time distribution types are presented in Appendix B at the end of the theses for the 15 sample sources described in Part I. The graph are based on simulation experiments.

These graphs exhibit the asymptotic properties set forward in the previous subsection. More research is needed to analyze the shape of the graphs.

I note here that great care must be taken when the generalized peakedness value of a process for a given holding time distribution is evaluated by means of simulation. The simulation must either start from the stable number of busy servers in the infinite server goup, or it may also start from zero busy servers but then the starting transient must be removed. I applied the latter method.

18 Application of generalized peakedness

Two main areas of application have been studied so far: approximation of loss systems and applications to delay systems. The first topic is described briefly and we concentrate on the second topic in more detail.

18.1 Approximations of loss systems

It is quite natural to apply generalized peakedness in the analysis of $G/G/s$ loss systems [13]. For such systems, we can characterize the arrival stream by its intensity and peakedness value with respect to the holding time distribution on the s-server group. To approximate blocking we may try to find a simple arrival process with the same intensity and peakedness as the original process for which the blocking can be calculated more easily.

18.2 Application to delay systems

In this case the main problem is to approximate the waiting time distribution of a $G/G/1/\infty$ FCFS queue. Contrary to the loss system approximation of the previous subsection, the choice of the appropriate holding time distribution where the peakedness value is taken is not at all obvious. The holding time should correspond to the "reaction time" of the system, which is the waiting time in our case. But it is just this that we want to approximate.

It is shown [14] that the appropriate choice for F (the complementary distribution function of the hypothetical infinite server group in the peakedness concept) is

$$F(x) = P(W > x | W > 0)$$

where W is the waiting time in the queue. The distribution of W is what we want to approximate. So if only a single peakedness value is available, it should be taken with respect to the best a priori estimate of the conditional delay distribution (such as a design objective); if several peakedness values are available, they should be with respect to a set of complementary distributions that span likely delay distributions.

Note that in spite of the fact that we are now dealing with a $G/G/1$ queue, we can still talk about the generalized peakedness of an arrival stream by offering the stream to an imaginary infinite server group.

18.3 Application of peakedness-based delay approximation in ATM: the $C_2/D/1$ queue

In this section the application of the concepts described in [14] is presented for an ATM-related topic. I consider a queue with deterministic service. The arrival stream is described by a Coxian-2 process. It is a renewal process which is able to cover both Erlang-2 ('smooth') and hyper-exponential ('bursty') distributions. There is one server and an infinite queue (buffer) with FCFS discipline. The problem is to approximate the waiting time distribution.

An approximation is that the waiting time W is exponentially distributed: [16]

$$P(W > x) = \alpha e^{-\beta x} \text{ for } x \geq 0 \ , \tag{1}$$

that is, for the waiting time W,

$$\left. \begin{array}{rcl} P(W = 0) & = & 1 - \alpha \\ P(W > 0) & = & \alpha \end{array} \right\} \tag{2}$$

and $W | W > 0$ is exponential with parameter β. The goal is to find α and β.

From this we have

$$\begin{aligned} \mathrm{E}(W) &= \alpha / \beta \\ &= \tau_r \mathrm{E}(S_a) + \tau_1 \mathrm{E}(Q_a) \\ &= \tau_r P(W > 0) + \tau_1 \mathrm{E}(Q_d) \end{aligned} \tag{3}$$

where S_a and Q_a denote the number of jobs in service (0 or 1) and in queue, respectively, just prior to an arbitrary arrival, Q_d denotes the number of jobs left in queue by an arbitrary job departing the queue (including those that do not wait in the queue) τ_1 is the mean service time of an arbitrary job, and τ_r is the mean remaining service time for the job in service at an arbitrary arrival if $S_a = 1$. The last equation follows by the equivalence of the queue distributions at arrivals to and departures from, the queue.

Q_d is just the number of arrivals to the queue during the waiting time of an arbitrary arrival. In our case (renewal arrival process) the length of the queue at a departure conditioned on the waiting time of the job departing is

$$\mathrm{E}(Q_d|W) = U(W) \tag{4}$$

(where $U(W)$ is the expectation function) because in the case of a renewal process, the number of arrivals in a time interval of length W after an arbitrarily chosen arrival does not depend on the past and therefore does not depend on W. (In cases where the arrival process is not renewal, Formula 4 would only be an approximation.)

From this,

$$\begin{aligned}
\mathrm{E}(Q_d) &= \mathrm{E}(\mathrm{E}(Q_d|W)) \\
&= \mathrm{E}(U(W)) \\
&= \int_0^\infty U(x) f_W(x) dx \\
&= \lim_{L\to\infty} \int_0^L U(x) f_W(x) dx \\
&= \lim_{L\to\infty} \left([U(x)F_W(X)]_0^L - \int_0^L \frac{dU(x)}{dx} F_W(x) dx \right) \\
&= \lim_{L\to\infty} \left(U(L)F_W(L) - \int_0^L dU(x) \right) + \lim_{L\to\infty} \int_0^L P(W > x) dU(x) \\
&= \lim_{L\to\infty} U(L)(F_W(L) - 1) + \int_0^\infty \alpha e^{-\beta x} dU(x) \\
&= -\alpha \lim_{L\to\infty} U(L) e^{-\beta L} + \alpha \int_0^\infty e^{-\beta x} dU(x) \\
&= \alpha \widetilde{U}(\beta)
\end{aligned} \tag{5}$$

where $f_W(x)$ and $F_W(x)$ are the probability density and distribution functions of W, and in the last step we suppose that $U(L)$ tends to infinity more slowly than exponentially.

Using this result in Equation 3 and making use of Equations 2, we get

$$\alpha/\beta = \tau_r \alpha + \tau_1 \alpha \widetilde{U}(\beta)$$

and for β,

$$\beta = \frac{1}{\tau_r + \tau_1 \widetilde{U}(\beta)}. \tag{6}$$

To connect $\widetilde{U}(\beta)$ to the peakedness for exponential holding times (since we suppose exponential waiting time, we take exponential service for the peakedness values, see Subsection 18.2), let us use Equation 6 in $z_{\exp}(\beta) = 1 + \widetilde{U}(\beta) - m/\beta$ of Subsection 17.3 to get

$$z_{\exp}(\beta) = 1 + \widetilde{U}(\beta)(1 - m\tau_1) - m\tau_r. \tag{7}$$

We can now express $\widetilde{U}(\beta)$ in terms of $z_{\exp}(\beta)$ and substitute it back to Equation 6

$$\beta = \frac{1/\tau_1 - m}{\tau_r/\tau_1 + z_{\exp}(\beta) - 1} \tag{8}$$

which gives us an implicit equation for β.

In our case of deterministic service, we may assume $\tau_r/\tau_1 = 1/2$ (which is accurate at high service utilization). $1/\tau_1$ is the service rate μ and so we have

$$\beta = \frac{\mu - m}{z_{\exp}(\beta) - 1/2}. \tag{9}$$

In our case of Coxian-2 arrival stream $\widetilde{U}(\beta)$ and then $z_{\exp}(\beta)$ may be evaluated in the following way. Let us denote the probability density function of the inter-arrival time distribution of the renewal process by $r(x)$ (which is a Coxian-2 distribution in this case).

The pdf of the length of 2 consecutive inter-arrivals is then $r(x) * r(x)$, of 3 consecutive inter-arrivals is $r(x) * r(x) * r(x)$ etc. It follows that the expected number of arrivals in an interval of dx and after x time units of an arrival is $r(x)dx + r(x) * r(x)dx + r(x) * r(x) * r(x)dx \ldots$ and so the expectation function is

$$U(t) = \int_0^t \Big(r(x) + r(x) * r(x) + r(x) * r(x) * r(x) + \ldots\Big) dx.$$

Denoting by $r^*(s)$ and $U^*(s)$ the Laplace transform of $r(x)$ and $U(t)$, respectively, we get

$$U^*(s) = \frac{1}{s}\left(\frac{r^*(s)}{1 - r^*(s)}\right)$$

and for the Laplace-Stieltjes transform of $U(t)$,

$$\widetilde{U}(s) = \frac{r^*(s)}{1 - r^*(s)}. \tag{10}$$

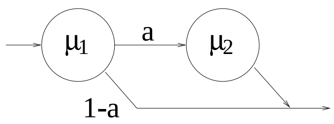


Figure 1: Coxian-2 inter-arrival time.

In our case, the inter-arrival time distribution is Coxian-2, which has three parameters (see Figure 1): μ_1, μ_2, a. (The inter-arrival distribution has two phases: phase 1 is exponential with parameter μ_1, phase 2 is zero with probability $1 - a$ and exponential with parameter μ_2 with probability a.)

The Laplace transform of the inter-arrival distribution is easily written as

$$r^*(s) = (1 - a)\left(\frac{\mu_1}{s + \mu_1}\right) + a\left(\frac{\mu_1}{s + \mu_1}\right)\left(\frac{\mu_2}{s + \mu_2}\right) = \frac{(1 - a)\mu_1 s + \mu_1\mu_2}{(s + \mu_1)(s + \mu_2)}.$$

This can be substituted into Equation 10 to get $\widetilde{U}(\beta)$ and by Equation 7 we get, after some algebra,

$$z_{\exp}(\beta) = 1 + ((1 - a)\mu_1 - m)\,\frac{1}{\beta + \mu_2 + a\mu_1} \tag{11}$$

using that

$$m = \frac{1}{\frac{1}{\mu_1} + \frac{a}{\mu_2}}$$

. To further simplify the notation, introduce

$$\begin{aligned} C_1 &= (1-a)\mu_1 - m \\ C_2 &= \mu_2 + a\mu_1 \end{aligned}$$

which reduces Equation 11 to

$$z_{\exp}(\beta) = 1 + C_1 \frac{1}{\beta + C_2}.$$

Using this result in Equation 9 gives a quadratic equation for determining β:

$$\frac{1}{2}\beta^2 + (C_1 + C_2/2 + m - \mu)\beta + C_2(m - \mu) = 0 \tag{12}$$

We can observe that the multiplication of the two roots of the above equation is

$$2C_2(m - \mu)$$

which is negative since $C_2 > 0$ and

$$m < \mu$$

otherwise we do not have a stable queue. It follows that the solution of Equation 12 for $\beta > 0$ is unique.

It remains to determine α. Denoting by N_b the number of jobs served in a busy period,

$$\alpha = P(W > 0) = 1 - \frac{1}{\mathrm{E}(N_b)}. \tag{13}$$

In the case of renewal arrival processes, for a busy period of length B, we may write

$$\mathrm{E}(N_b|B) = 1 + U(B)$$

because the busy period begins with the service of the first job and there are $U(B)$ other jobs expected to come in the busy period.

Assuming that the the length of a busy period, B, is exponentially distributed with parameter γ, we can derive similarly to Equation 5 that

$$\mathrm{E}(N_b) = 1 + \widetilde{U}(\gamma). \tag{14}$$

Using the approximating Equation 14 in Equation 13 gives

$$\alpha = 1 - \frac{1}{1 + \widetilde{U}(\gamma)}.$$

Motivated by the $GI/M/1$ and $M/G/1$ queues, an approximation for γ is [13]

$$\gamma \approx \beta\tau_r/\tau_1 \approx \beta/2. \tag{15}$$

By $z_{\exp}(\gamma) = 1 + \widetilde{U}(\gamma) - m/\gamma$ of Subsection 17.3 and the approximation in Equation 15 for γ gives

$$z_{\exp}(\beta/2) + 2m/\beta = 1 + \widetilde{U}(\gamma)$$

and finally

$$\alpha = 1 - \frac{1}{z_{\exp}(\beta/2) + 2m/\beta} \tag{16}$$

where $z_{\exp}(\beta/2)$ is given by Equation 11.

18.4 Simulation of $C_2/D/1$ queue

In order to assess the accuracy of the approximation method set forward in the previous subsection, the method must be compared with either the exact solution, other approximation methods or simulation results.

It is possible to give the exact formulae for the waiting time distribution of the $C_2/D/1$ queue, but it is not very useful for practical computations [17]. In [17] another approximation is shown, but that also uses the exponential assumption, and requires a very long numerical computation.

Therefore I have built a simulation model to compare it to the generalized peakedness based approximation.

The simulation was carried out in the following way. The input parameters were the parameters of the Coxian-2 arrival process and the deterministic service time of the server given in terms of the utilization. The $C_2/D/1$ queue was simulated, and the output was the waiting time of the all the arriving cells.

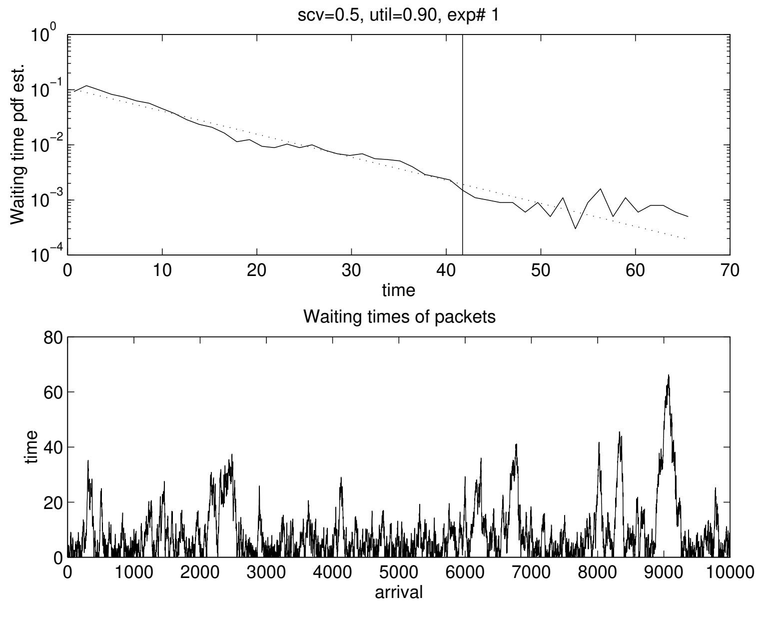


Figure 2: Results of a simulation of the waiting time distribution. (The simulation length was 10000 arrivals, Erlang-2 arrival process with mean 2 and 90% server utilization.)

Figure 2 shows an example of the simulation output. Figure 2, bottom shows the waiting times of the individual cells. (The simulation length was 10000 arrivals, Erlang-2 arrival process with mean 2 and 90% server utilization.) It must be noted that as the server utilization goes to 1, the queue becomes less and less stable, and there will be long-term fluctuations with long periods of high waiting time. On the figure we can notice one such period near the 9000th arrival. Since these periods are rare, we must have an extremely long simulation to take into account the long waiting times caused by these periods. For this reason, utilization was never set to more than 0.95 to avoid nearly instable systems. To eliminate the effect of the rare long waiting times, I did not consider waiting times longer than a certain limit, which was 42 in this case. (In general, I did not consider waiting times above the limit where less than 20 samples are available.)

First, α was estimated from the set of waiting time values as the proportion of zero waiting times in the whole data set.

Figure 2, top shows the logarithm of the estimated waiting time density function for positive waiting times. The vertical line shows the limit (42 in this case) above which the values were not taken into account. If our assumption is true and the positive waiting times are distributed according to $\beta e^{-\beta x}$, then the logarithm plot should be as straight descending line. Therefore a least squares regression line was fitted to the data shown by the dotted line. It can be seen that the exponential waiting time distribution is not far from reality. The negative slope of the fitted regression line is the estimation for β.

I note here that although $\beta e^{-\beta x}$ should fit well to the data, the dotted line is not this, but $K e^{-\beta x}$, where K is a constant different from β. The value of K is irrevelant and is influenced by the distribution above the limit (the vertical line).

18.5 Comparison of generalized peakedness based waiting time approximation with simulation results

I compared the waiting time distribution parameters (α and β) approximated using the generalized peakedness concept (α_{gp} and β_{gp}) with the estimations from the simulations (α_{sm} and β_{sm}). The comparison was done for four data sets of the Coxian-2 inter-arrival distribution: Erlang-2, Poisson and two Hyper-exponential arrivals with squared coefficients of variation 0.5, 1, 1.96, 5.54 respectively. In each case, four server utilization values (and the corresponding service time) were used: 0.95, 0.9, 0.8, 0.5. The simulation consisted of 10000 arrivals and was repeated 10 times in order to establish 90% confidence intervals.

The results show that the accuracy of the generalized peakedness method is greater at higher values of squared coefficient of variation. For the Erlang-2 arrival process the method is inaccurate. In the case of the Poisson arrival process, the approximation gives exact results. In the case of the first hyper-exponential arrival processes, the accuracy is high for high utilization and decreases as the utilization decreases. In the case of the second hyper-exponential arrival process with squared coefficient of variation 5.54, the method is accurate.

Erlang-2

ρ	α_{gp}	α_{sm}	β_{gp}	β_{sm}
0.95	0.9066	0.9498 ± 0.0140	0.1005	0.0563 ± 0.0228
0.90	0.8239	0.9009 ± 0.0148	0.2034	0.1148 ± 0.0273
0.80	0.6800	0.7966 ± 0.0098	0.4254	0.2653 ± 0.0331
0.50	0.3431	0.4985 ± 0.0145	1.4142	1.0974 ± 0.1013

Table 2: Waiting time distribution approximation $\alpha e^{-\beta x}$. The first column is the server utilization, the second and fourth are the approximations based on generalized peakedness, the third and fifth are the approximations based on simulation. The arrival process is *Erlang-2* (Coxian-2 arrival process parameters: $a = 1$, $\mu_1 = 1$, $\mu_2 = 1$); mean inter-arrival time is 2, squared coefficient of variation is 0.5.

Poisson

ρ	α_{gp}	α_{sm}	β_{gp}	β_{sm}
0.95	0.9500	0.9481 ± 0.0074	0.1053	0.1138 ± 0.0428
0.90	0.9000	0.8941 ± 0.0164	0.2222	0.2297 ± 0.0671
0.80	0.8000	0.7991 ± 0.0120	0.5000	0.5251 ± 0.0677
0.50	0.5000	0.4998 ± 0.0105	2.0000	2.1692 ± 0.1694

Table 3: Waiting time distribution approximation $\alpha e^{-\beta x}$. The first column is the server utilization, the second and fourth are the approximations based on generalized peakedness, the third and fifth are the approximations based on simulation. The arrival process is *Poisson* (Coxian-2 arrival process parameters: $a = 0$, $\mu_1 = 1$, $\mu_2 = 1$); mean inter-arrival time is 1, squared coefficient of variation is 1.

19 Concluding remarks on generalized peakedness

In this part the generalized peakedness of teletraffic processes was defined as the variance-to-mean ratio of the number of busy servers in an infinite group of servers to which the stationary arrival stream is hypothetically offered. It is a functional which maps complementary holding time distributions to peakedness values. The definition expresses the idea that the variability (i.e. burstiness) of the traffic in a system does not solely depend on the arrival process but is also determined by the holding times of the servers in the system.

In very special cases this functional can be formulated explicitly. However, usually it is not possible to do it analytically. The case of exponential holding times plays a very important role because the generalized peakedness values as a function of the mean of the exponential holding time provides a complete second order characteristic. The generalized peakedness value at exponential holding times are much easier to compute and can be used to express generalized peakedness values at other holding time distributions, such as Coxian.

Hyper-exponential (1.96)

ρ	α_{gp}	α_{sm}	β_{gp}	β_{sm}
0.95	0.9740	0.9717 ± 0.0104	0.0867	0.0732 ± 0.0471
0.90	0.9468	0.9425 ± 0.0111	0.1846	0.1732 ± 0.0447
0.80	0.8889	0.8854 ± 0.0114	0.4244	0.3884 ± 0.0737
0.50	0.6775	0.7065 ± 0.0062	1.8856	1.6516 ± 0.1086

Table 4: Waiting time distribution approximation $\alpha e^{-\beta x}$. The first column is the server utilization, the second and fourth are the approximations based on generalized peakedness, the third and fifth are the approximations based on simulation. The arrival process is *Hyper-exponential* (Coxian-2 arrival process parameters: $a = 0.5$, $\mu_1 = 8$, $\mu_2 = 1$); mean inter-arrival time is 0.625, squared coefficient of variation is 1.96.

Hyper-exponential (5.54)

ρ	α_{gp}	α_{sm}	β_{gp}	β_{sm}
0.9500	0.9907	0.9836 ± 0.0083	0.0679	0.0851 ± 0.0527
0.9000	0.9810	0.9693 ± 0.0107	0.1444	0.1722 ± 0.0742
0.8000	0.9596	0.9440 ± 0.0124	0.3303	0.3235 ± 0.0673
0.5000	0.8753	0.8633 ± 0.0116	1.4471	1.3194 ± 0.1497

Table 5: Waiting time distribution approximation $\alpha e^{-\beta x}$. The first column is the server utilization, the second and fourth are the approximations based on generalized peakedness, the third and fifth are the approximations based on simulation. The arrival process is *Hyper-exponential* (Coxian-2 arrival process parameters: $a = 0.25$, $\mu_1 = 32$, $\mu_2 = 1$); mean inter-arrival time is 0.2812, squared coefficient of variation is 5.54.

The concept of generalized peakedness can be applied to both loss and delay systems. In this theses an application was presented for the approximation of waiting time distribution in a $C_2/D/1$ queue. Possible application of the results include ATM network design. The importance of the approximation is that it is very simple (it leads to a quadratic equation) as opposed to the exact solution and other approximation method found in the literature. Simulation experiments show that this approximation method is accurate when the squared coefficient of variation of the arrival process is at least one and the server utilization is not too low.

The possible use of the generalized peakedness concept to teletraffic is promising and not yet elaborated. Further research areas include

- applying generalized peakedness for the estimation of cell loss probabilities and buffer sizes;
- investigating the scope of queue systems where the generalized peakedness concept gives a simple and accurate approximation;

- developing a framework for the exact and approximatie computation of generalized peakedness in important practical cases;
- investigating the possibility of measuring generalized peakedness of a process for exponential holding time distribution;
- relating the generalized peakedness values of teletraffic processes to other characterization methods such as IDI, IDC, leaky buckets.

Appendix

A Characteristics of traffic models

In the following pages, the result of applying the various traffic characterization methods to the 15 sample sources are shown. Each page is devoted to one traffic source. The simulation involved 10000 inter-arrival times for each sources. On top of the graphs, the name of the plot and the number of experiments from which the average was computed are shown. The accuracy of the estimations are indicated by the 90% confidence intervals.

On the top of the pages the first four moments of inter-arrival times are given, then the squared coefficient of variation for the inter-arrival times followed by the peak-to-mean ratio.

The figures are: a sample from the intensity-time plot, the estimations of the inter-arrival probability density function, the correlation of the sequence of inter-arrival times and of the sequence of counts of arrivals, the index of dispersion for intervals and counts, the variance-time curve and autocorrelation between two successive intervals of length t.

The following labels were used to identify the graphs:

`intensity` for the intensity-time plot,

`pdf` for the probability density function estimate of the inter-arrival length,

`correlation` for the correlation in the sequence of inter-arrival times,

`dcorrelation` for the correlation in the sequence of counts of arrivals in successive intervals of equal length,

`IDI` for the index of dispersion for intervals,

`IDC` for the index of dispersion for counts,

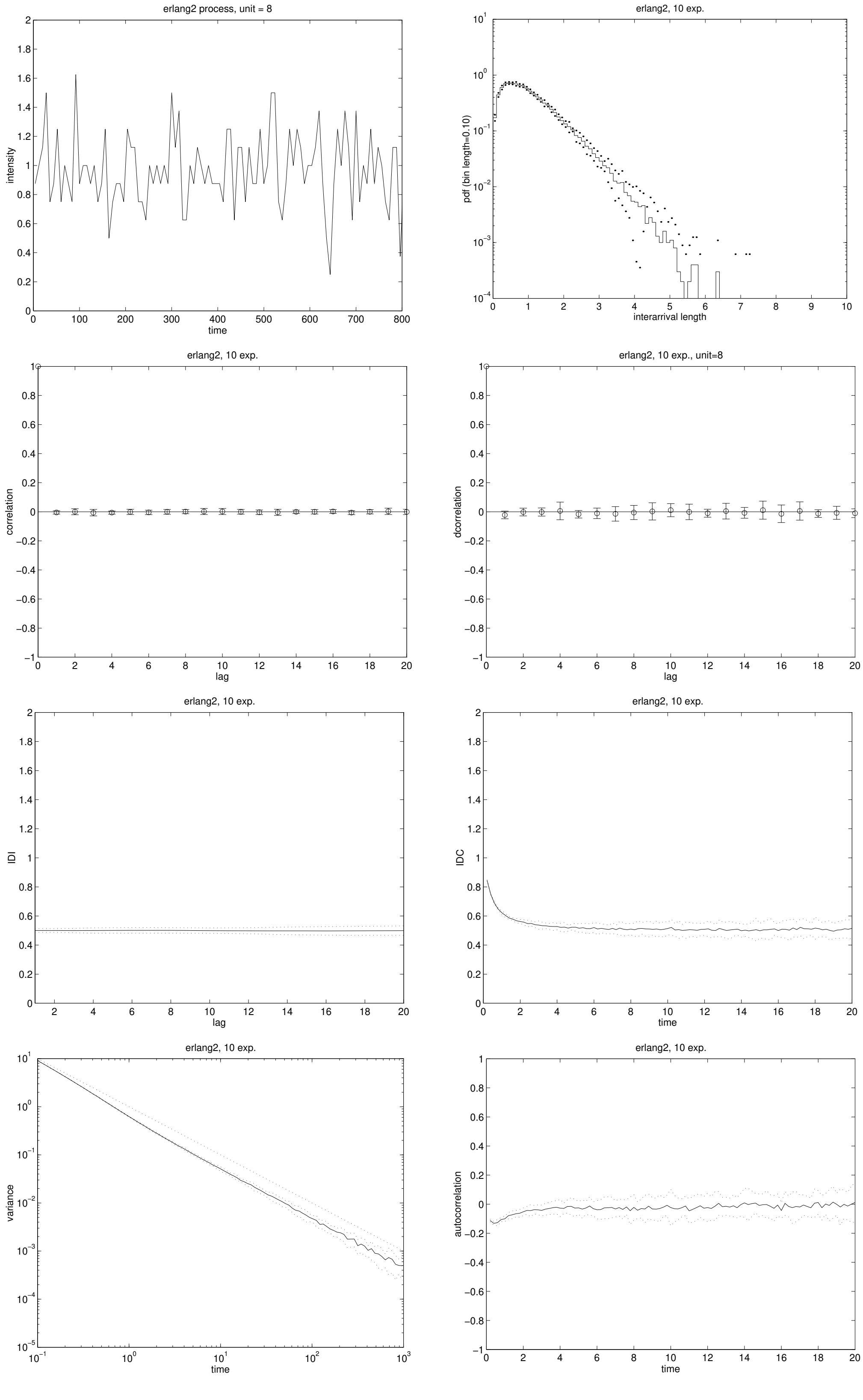
`vartime` for the variance-time plot, and

`autocorrelation` for the correlation in the number of arrivals in two successive intervals of length t.

See Part III and Section 13 for more details on the interpretation of the figures.

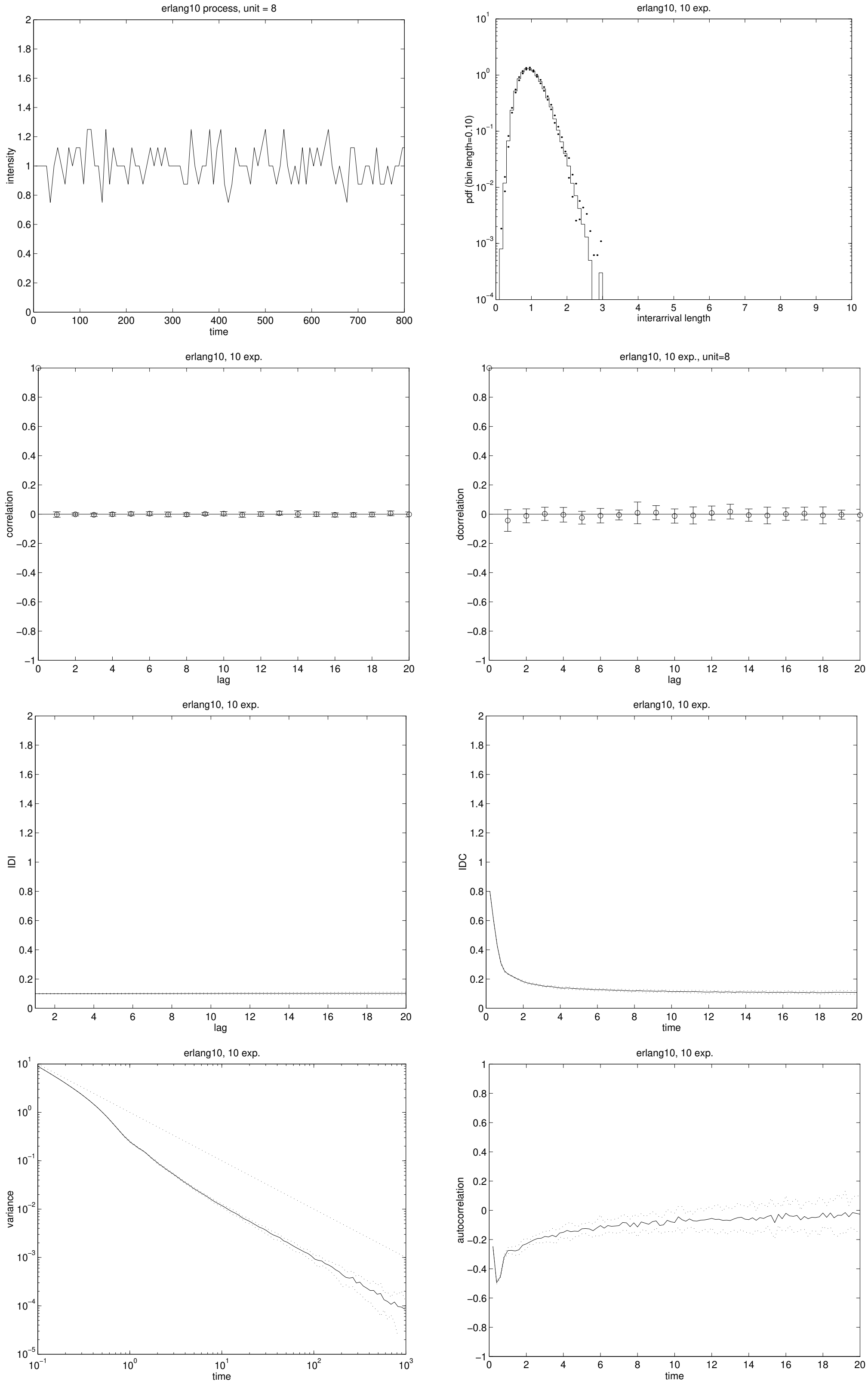
`erlang2`

$m_1 = 0.999 \pm 0.011$, $m_2 = 1.49 \pm 0.028$, $m_3 = 3.028 \pm 0.12$, $m_4 = 7.39 \pm 0.45$,
$C^2 = 0.500 \pm 0.014$, $P/M = 400 \pm 720$



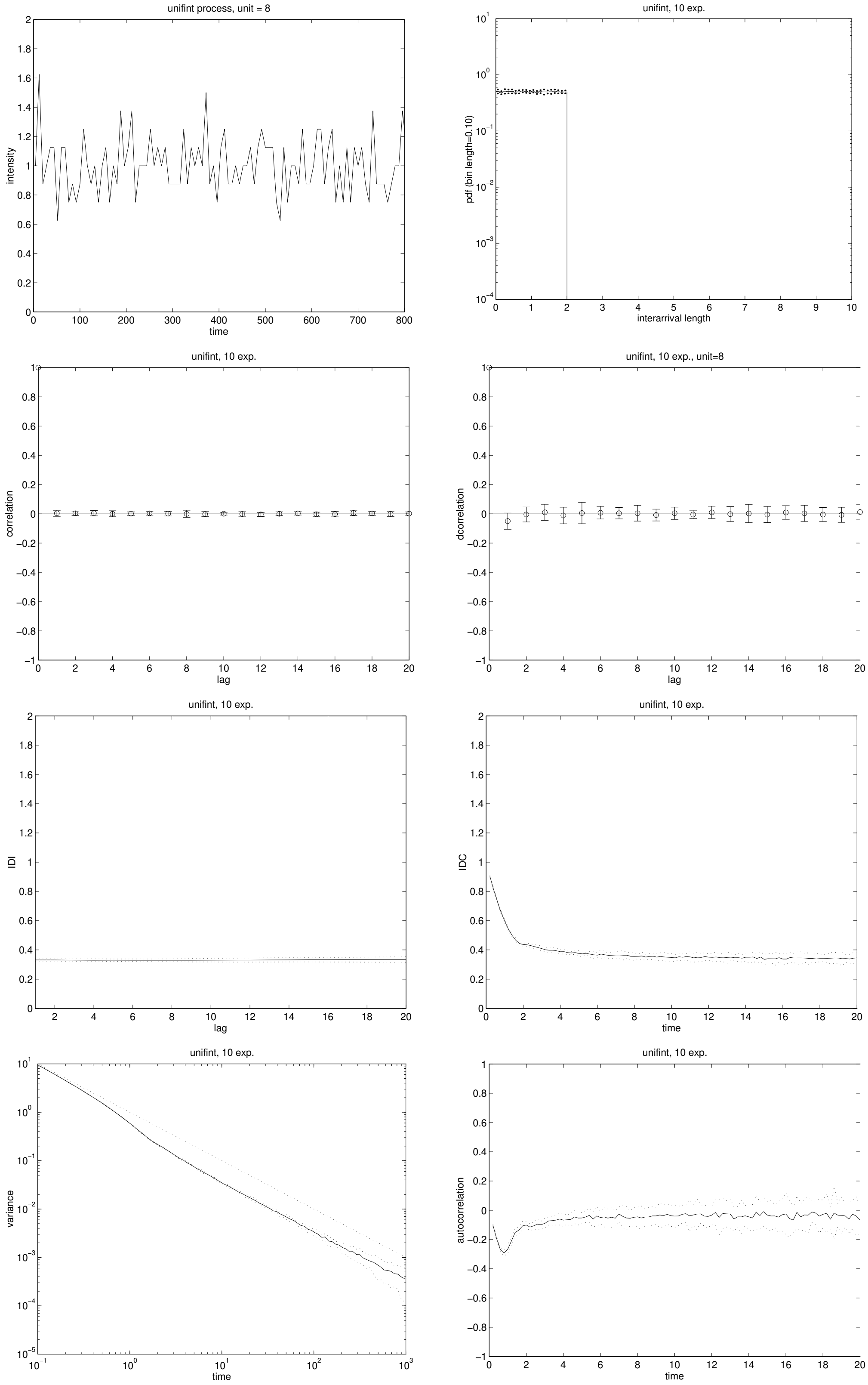
$\boxed{\texttt{erlang10}}$

$m_1 = 1.002 \pm 0.008$, $m_2 = 1.097 \pm 0.01$, $m_3 = 1.32 \pm 0.03$, $m_4 = 1.73 \pm 0.06$,
$C^2 = 0.1005 \pm 0.002$, $P/M = 5.3 \pm 1.8$



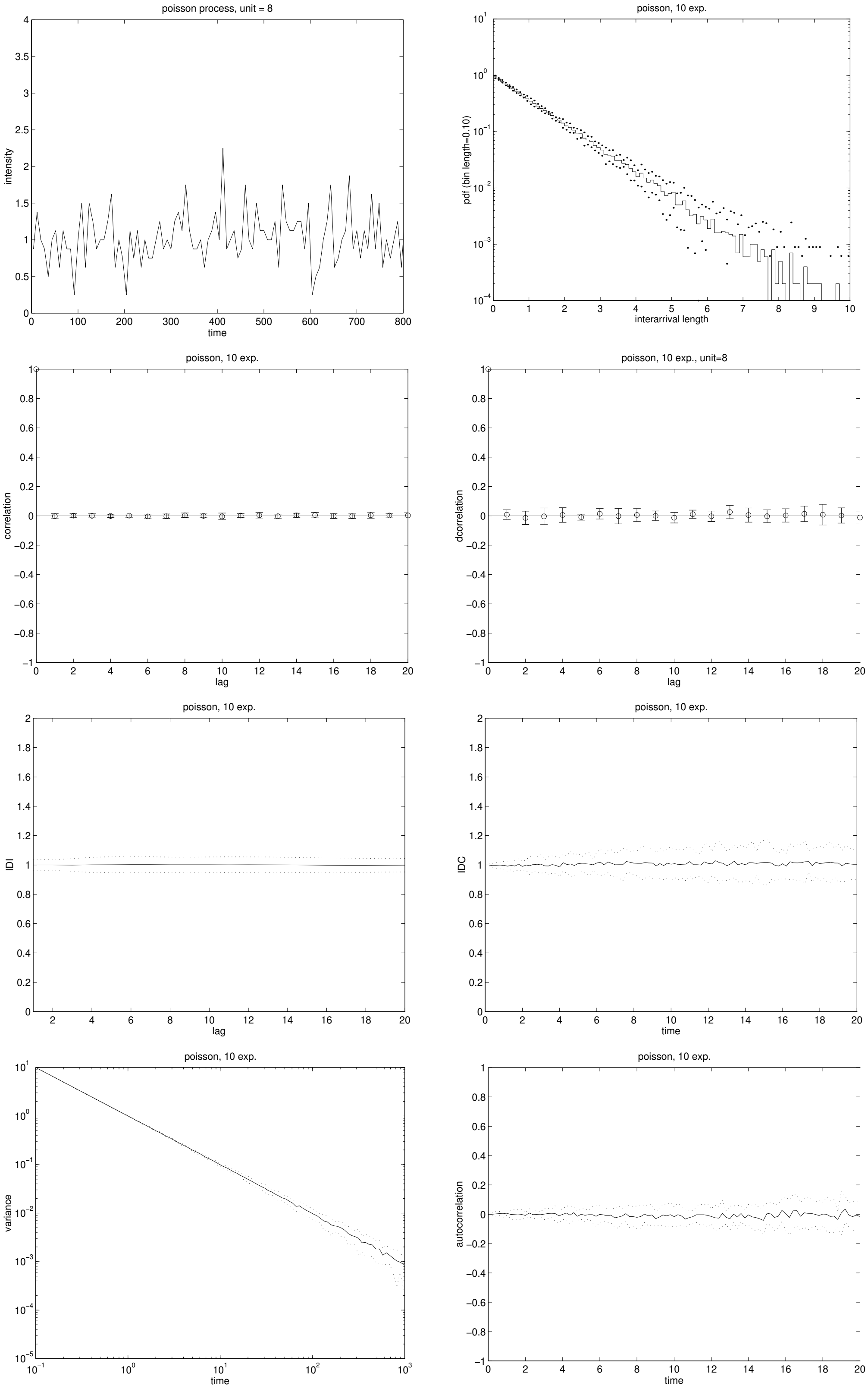
$\boxed{\texttt{unifint}}$

$m_1 = 1.000 \pm 0.008$, $m_2 = 1.33 \pm 0.02$, $m_3 = 1.99 \pm 0.03$, $m_4 = 3.19 \pm 0.08$,
$C^2 = 0.333 \pm 0.006$, $P/M = 20.5 \cdot 10^3 \pm 66 \cdot 10^3$



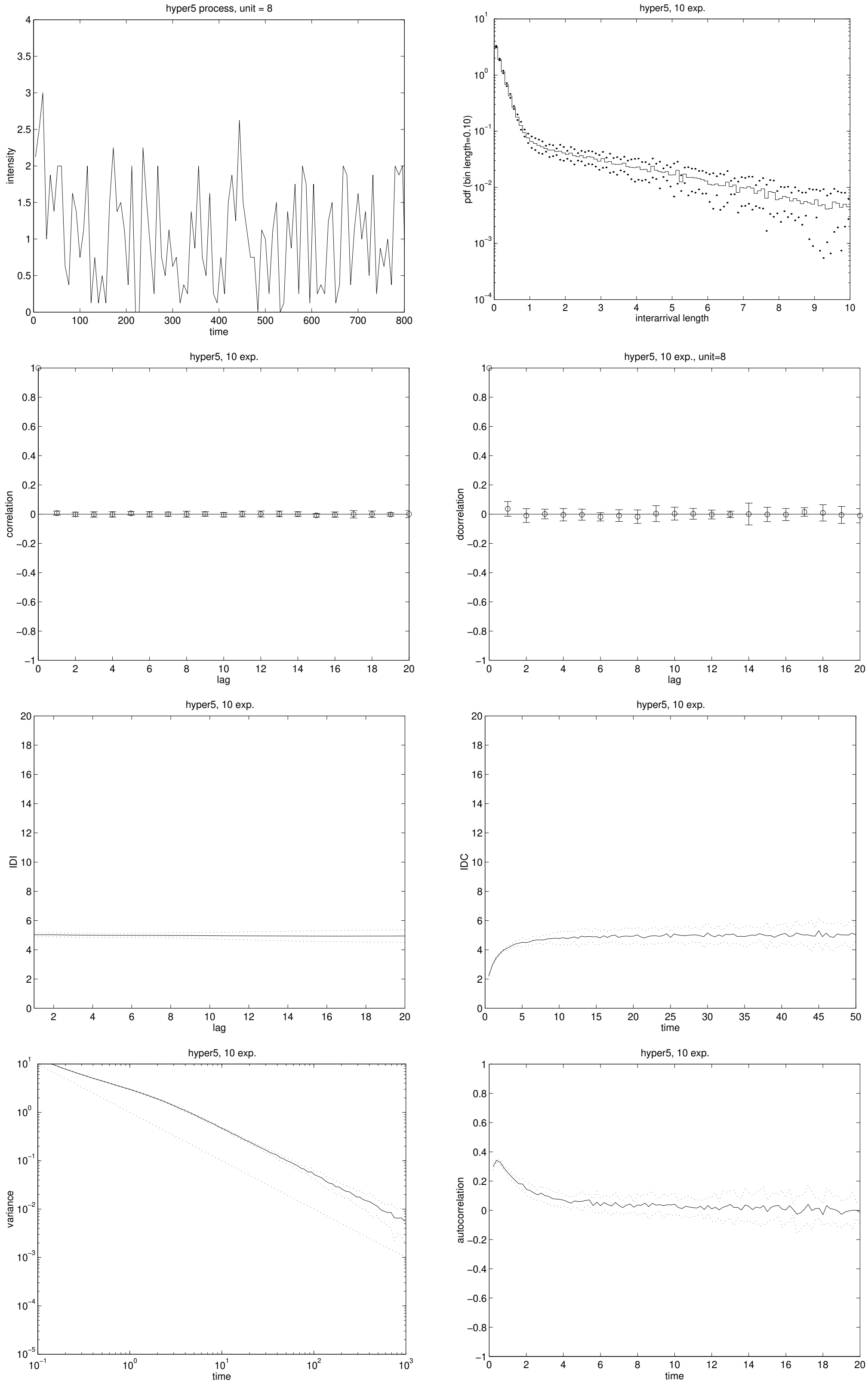
$\boxed{\texttt{poisson}}$

$m_1 = 0.998 \pm 0.012$, $m_2 = 1.997 \pm 0.087$, $m_3 = 6.02 \pm 0.42$, $m_4 = 25 \pm 4.7$, $C^2 = 1.001 \pm 0.03$, $P/M = 21.5 \cdot 10^3 \pm 60 \cdot 10^3$



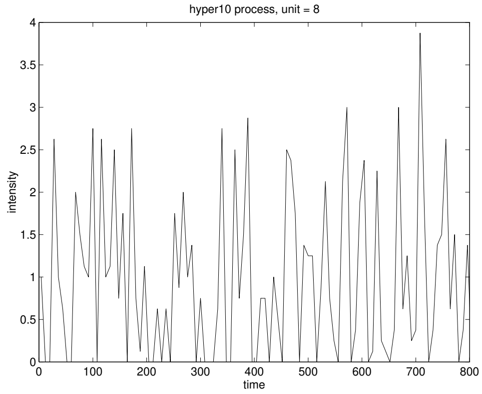
hyper5

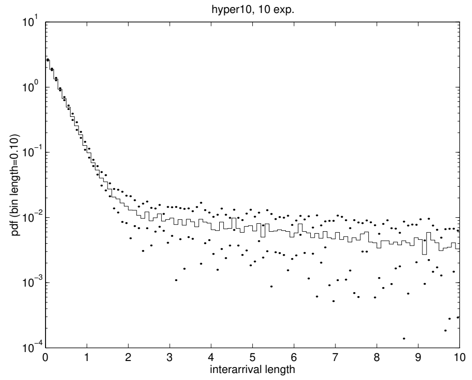
$m_1 = 1.000 \pm 0.03$, $m_2 = 5.9 \pm 0.43$, $m_3 = 62 \pm 14$, $m_4 = 850 \pm 208$,
$C^2 = 5.03 \pm 0.29$, $P/M = 95 \cdot 10^3 \pm 146 \cdot 10^3$

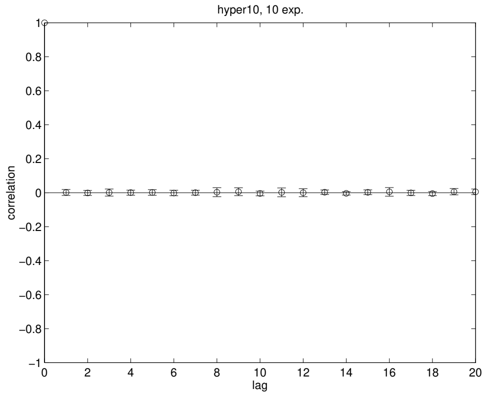


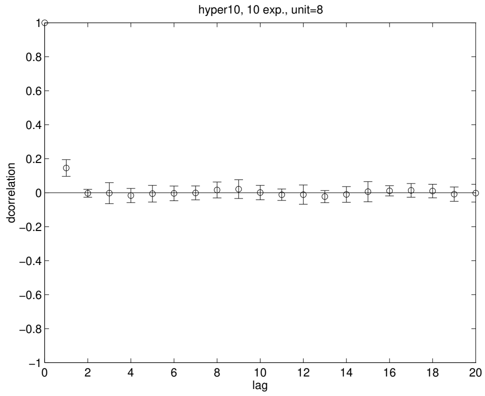
hyper10

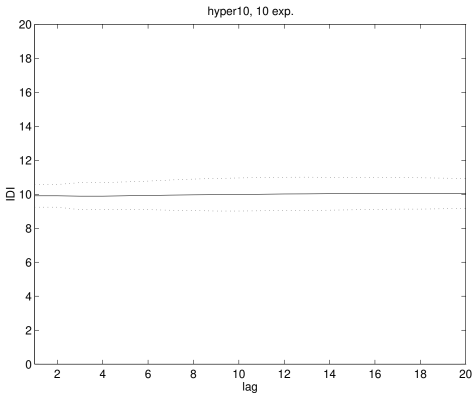
$m_1 = 1.008 \pm 0.05$, $m_2 = 11 \pm 0.96$, $m_3 = 238 \pm 49$, $m_4 = 7800 \pm 3350$,
$C^2 = 10.2 \pm 0.68$, $P/M = 113 \cdot 10^3 \pm 189 \cdot 10^3$

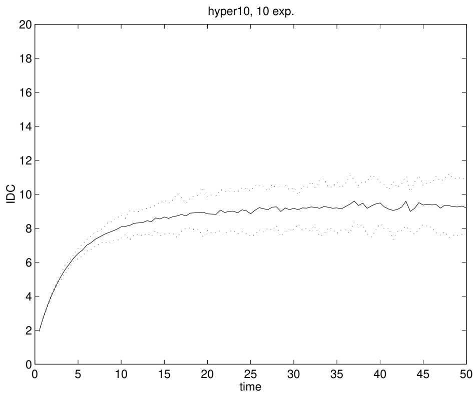


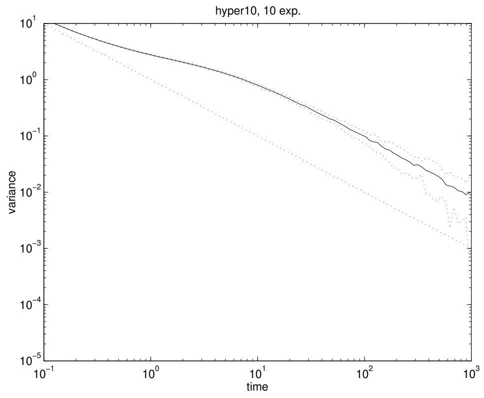


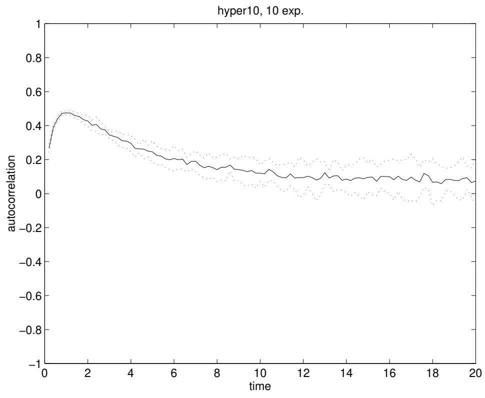






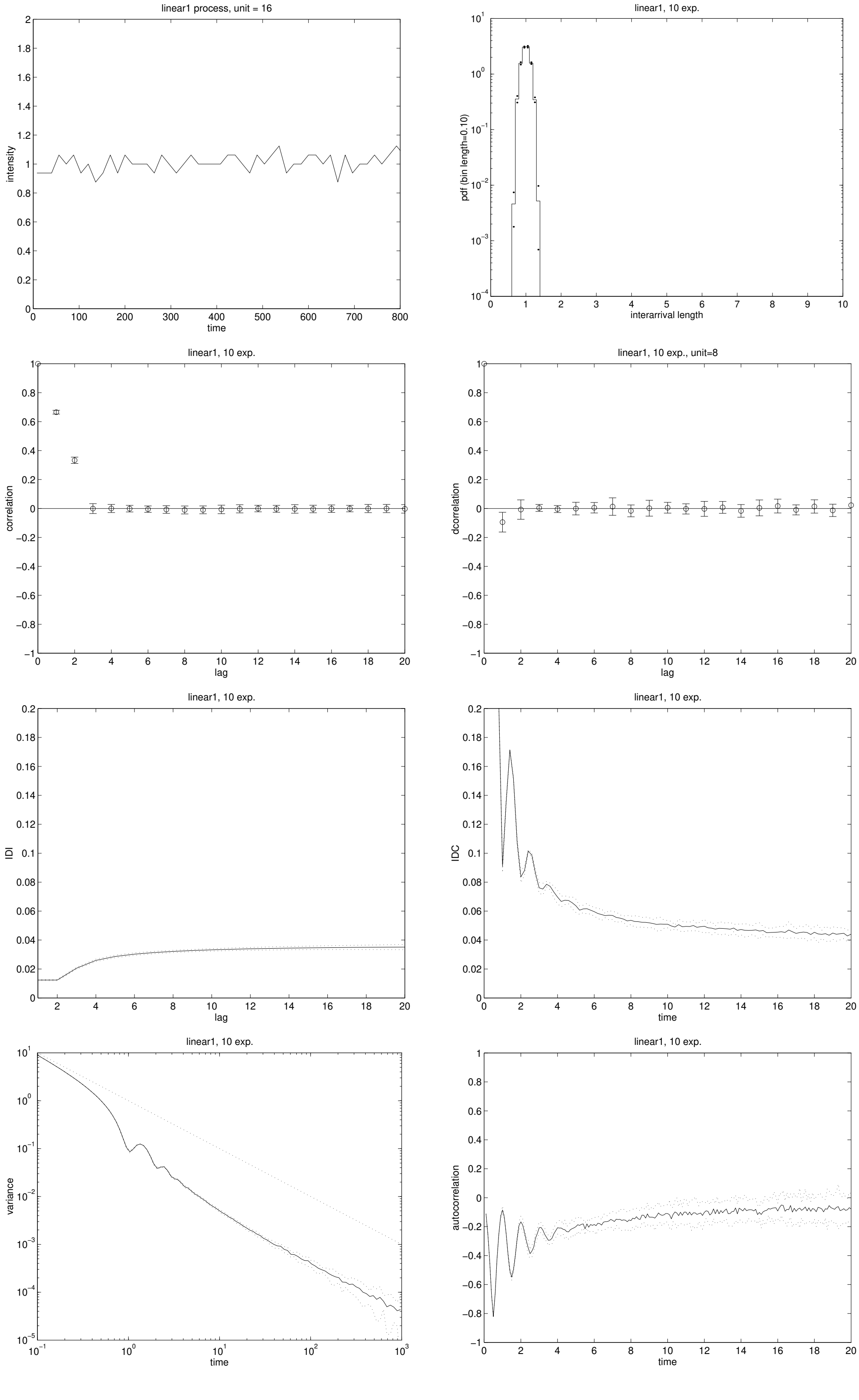






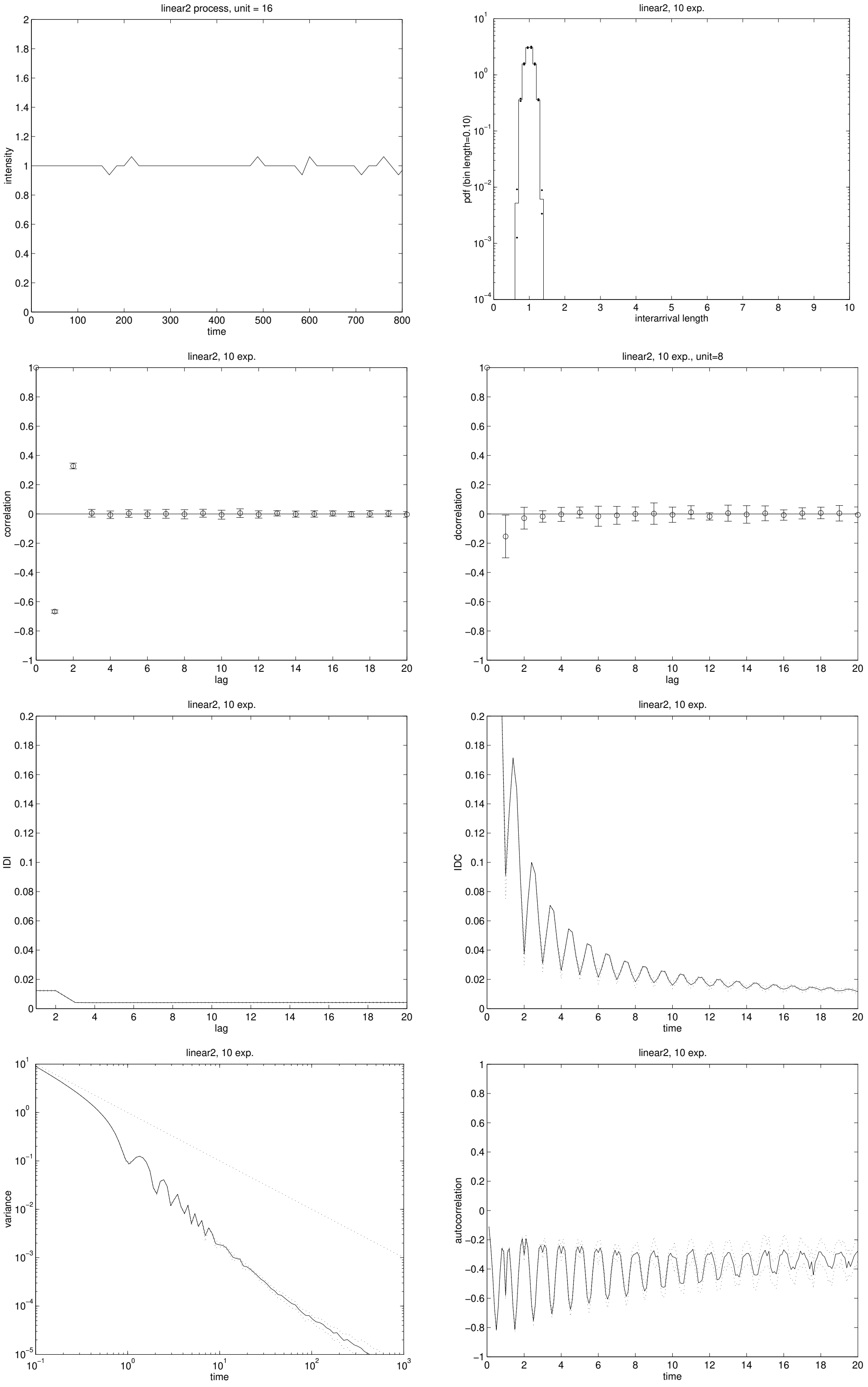
$\boxed{\texttt{linear1}}$

$m_1 = 1.0006 \pm 0.004$, $m_2 = 1.01 \pm 0.005$, $m_3 = 1.036 \pm 0.011$, $m_4 = 1.078 \pm 0.013$, $C^2 = 0.0123 \pm 0.0003$, $P/M = 1.46 \pm 0.021$



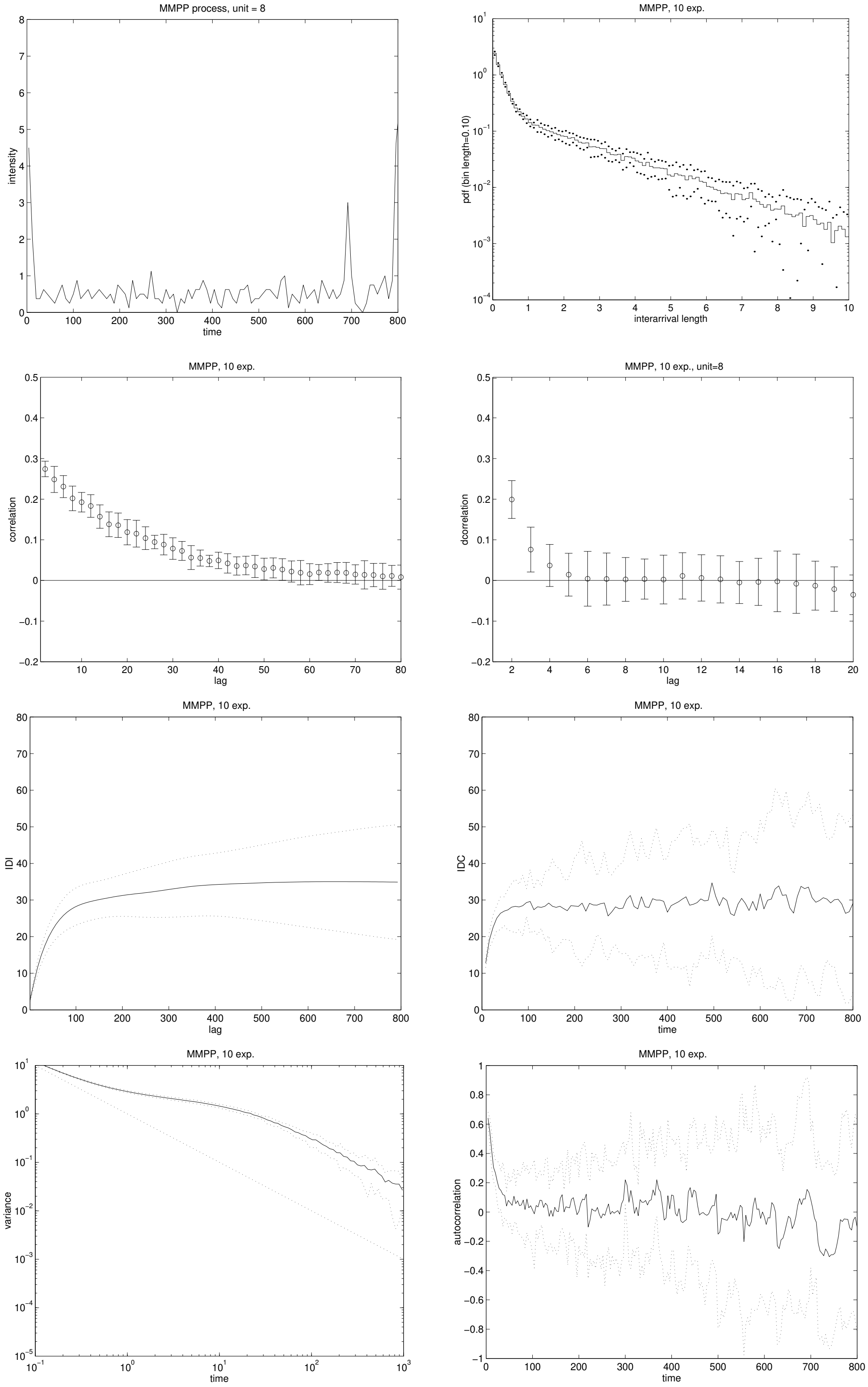
$\boxed{\texttt{linear2}}$

$m_1 = 0.9997 \pm 0.0006$, $m_2 = 1.0129 \pm 0.0015$, $m_3 = 1.0367 \pm 0.0027$, $m_4 = 1.074 \pm 0.006$, $C^2 = 0.0123 \pm 0.0003$, $P/M = 1.46 \pm 0.018$



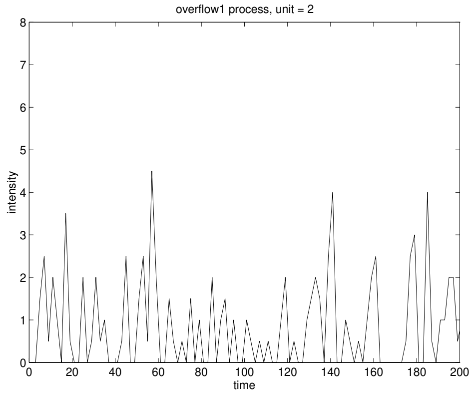
$\boxed{\texttt{MMPP}}$

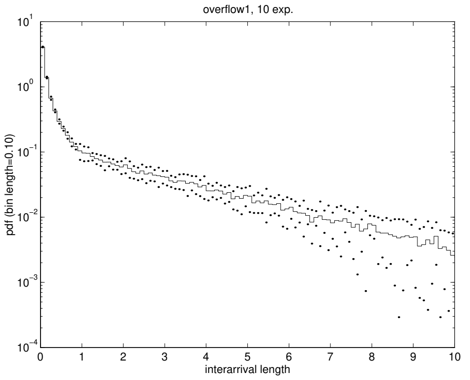
$m_1 = 1.012 \pm 0.07$, $m_2 = 3.9 \pm 0.42$, $m_3 = 23.025 \pm 3.3$, $m_4 = 177 \pm 31$,
$C^2 = 2.5 \pm 0.3$, $P/M = 84 \cdot 10^3 \pm 228 \cdot 10^3$

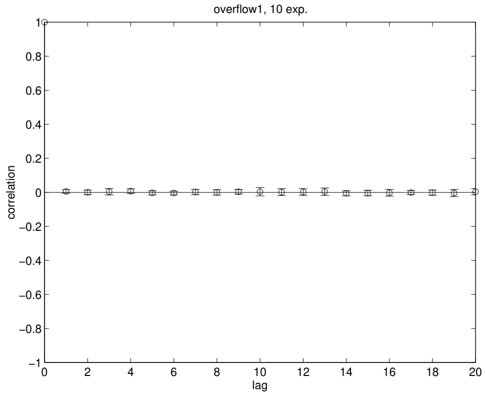


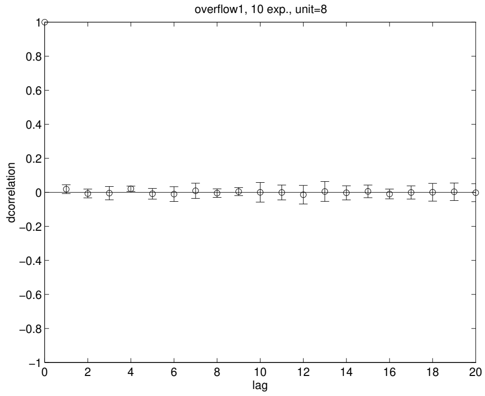
$\boxed{\texttt{overflow1}}$

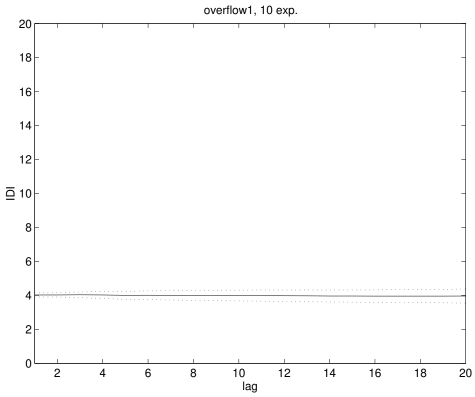
$m_1 = 1.0000 \pm 0.0000$, $m_2 = 4.96 \pm 0.17$, $m_3 = 40.6 \pm 1.8$, $m_4 = 448 \pm 94$, $C^2 = 4.03 \pm 0.17$, $P/M = 1.9 \cdot 10^6 \pm 5 \cdot 10^6$

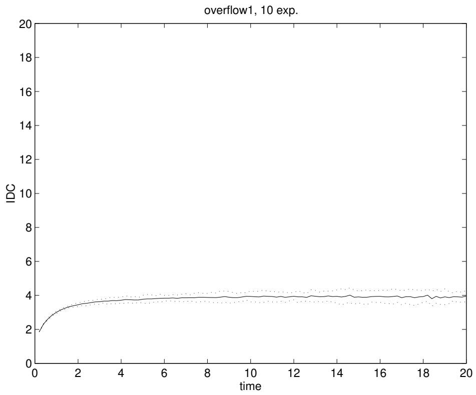


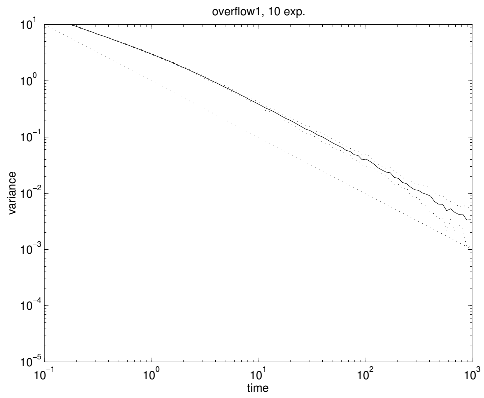


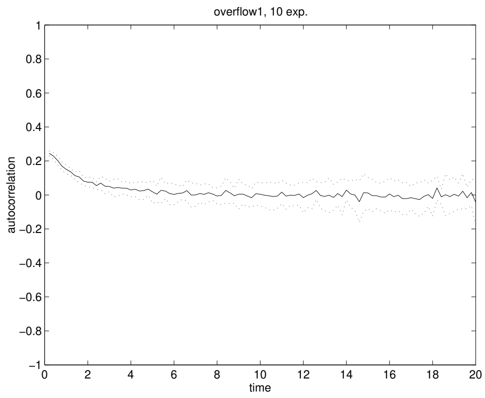






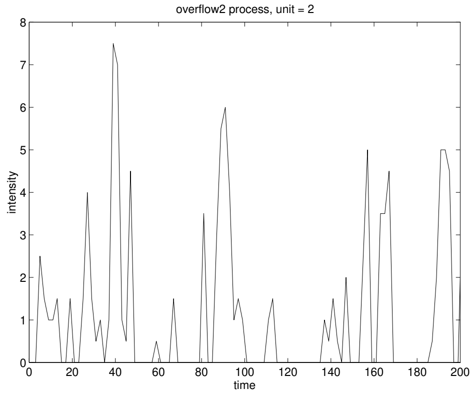


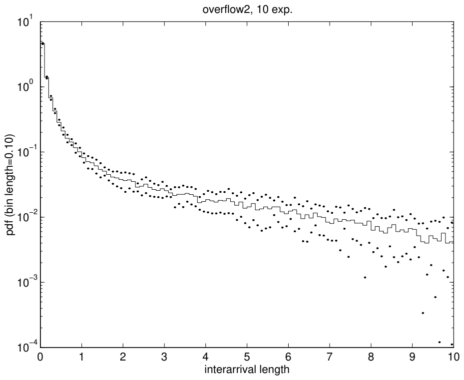


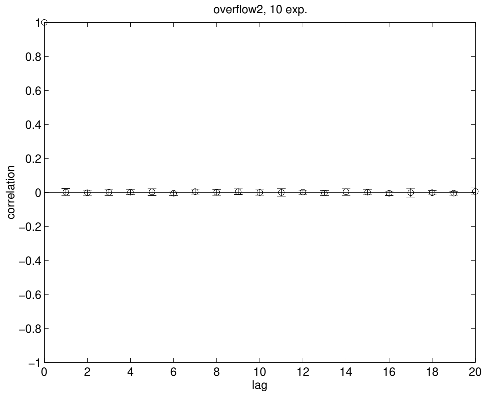


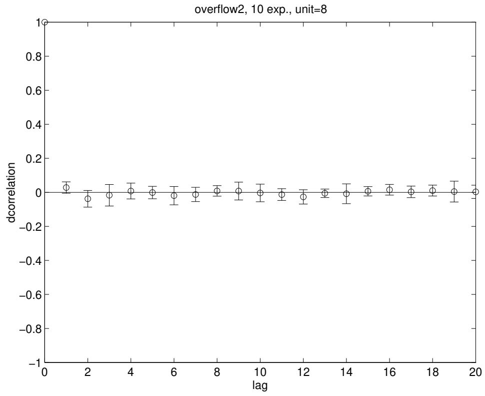
$\boxed{\texttt{overflow2}}$

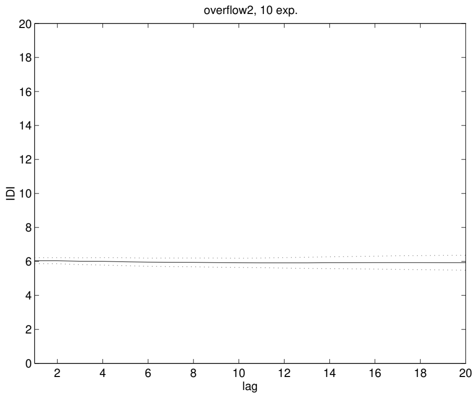
$m_1 = 1.0000 \pm 0.0000$, $m_2 = 7.03 \pm 0.33$, $m_3 = 82 \pm 13$, $m_4 = 1460 \pm 280$,
$C^2 = 6.08 \pm 0.26$, $P/M = 236 \cdot 10^3 \pm 487 \cdot 10^3$

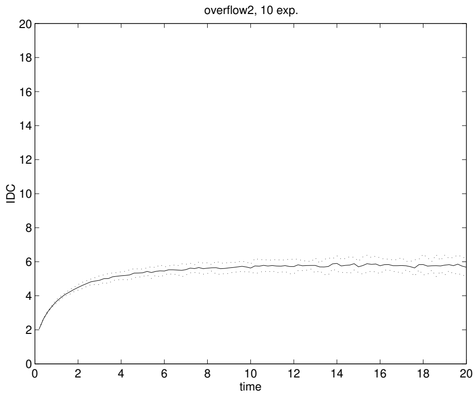


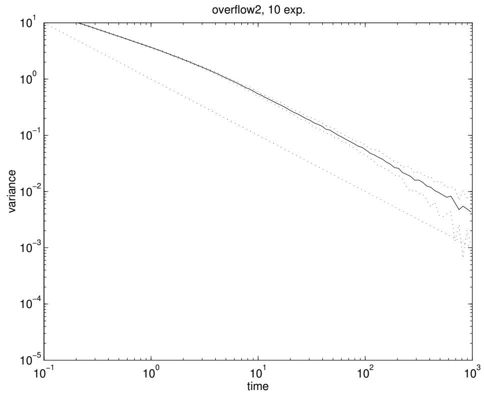


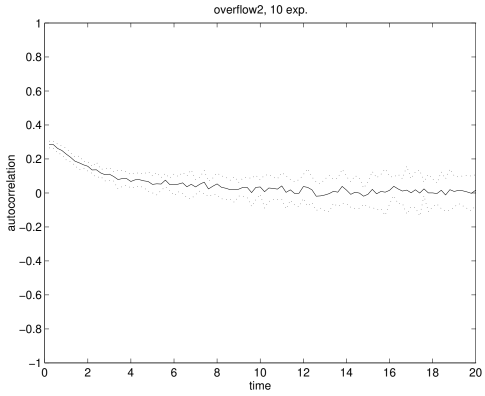






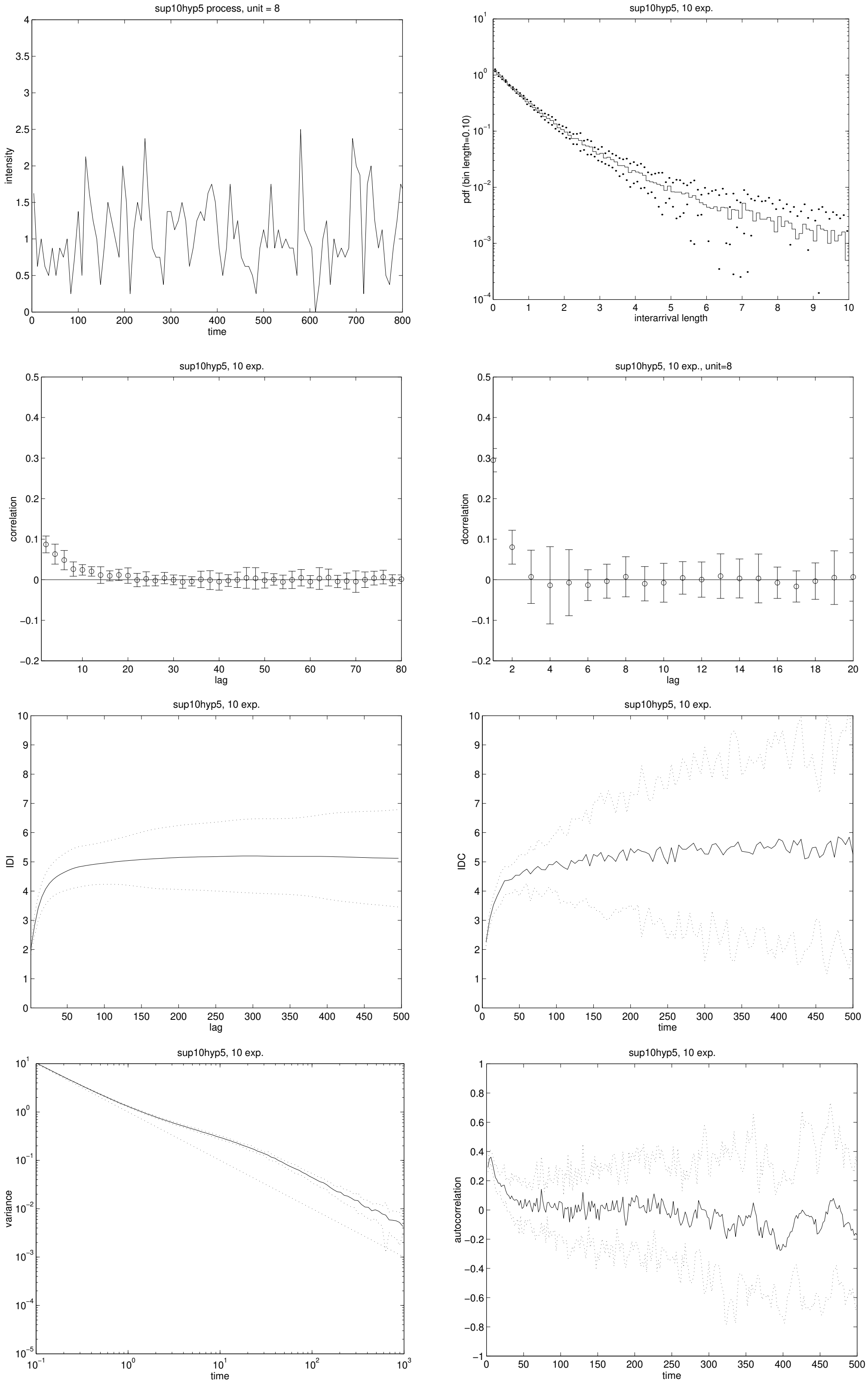






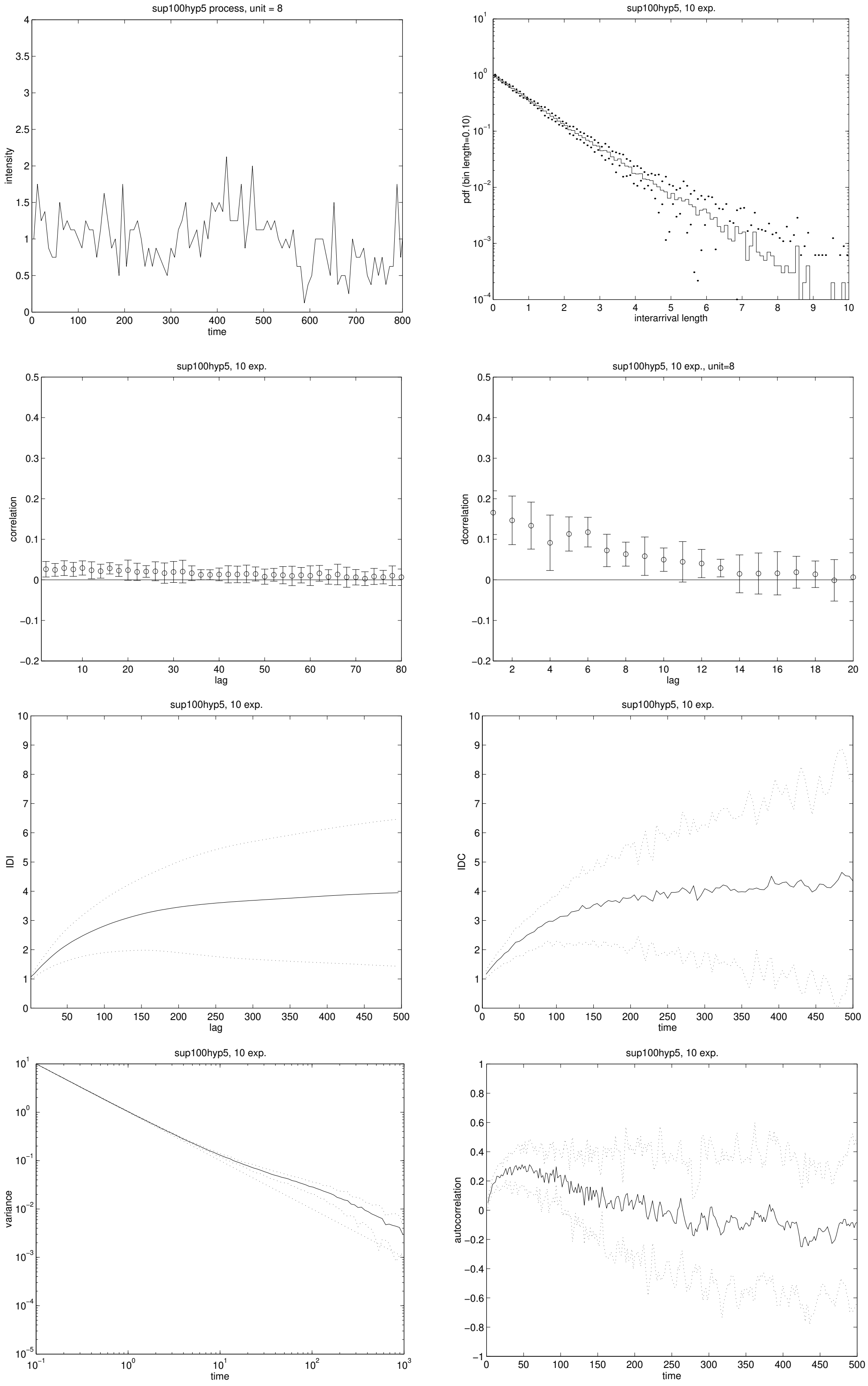
$\boxed{\texttt{sup10hyp5}}$

$m_1 = 1.008 \pm 0.05$, $m_2 = 3.00 \pm 0.36$, $m_3 = 22.4 \pm 5.7$, $m_4 = 290 \pm 168$,
$C^2 = 2.01 \pm 0.17$, $P/M = 16 \cdot 10^3 \pm 17 \cdot 10^3$



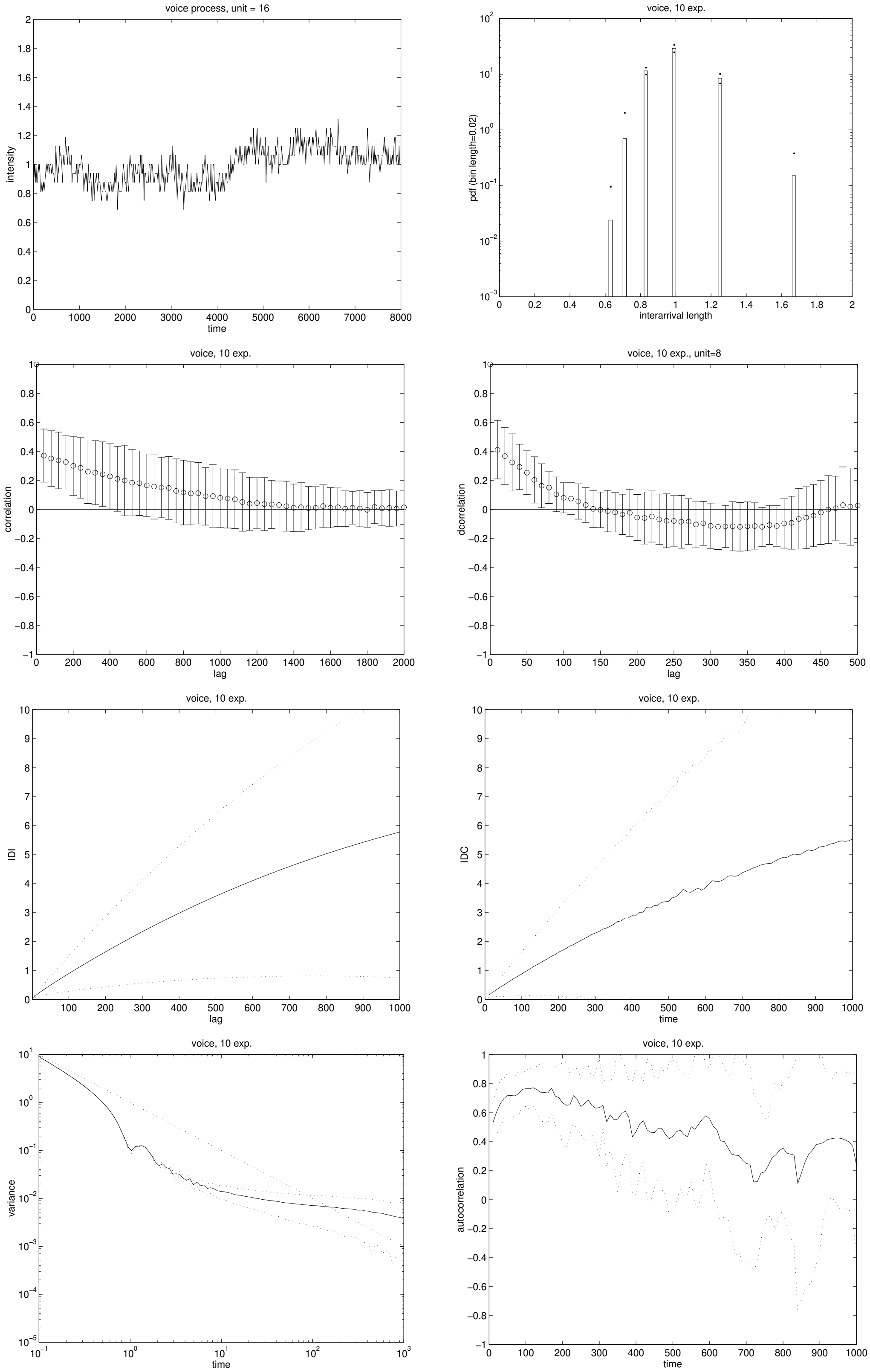
$\boxed{\texttt{sup100hyp5}}$

$m_1 = 1.0095 \pm 0.05$, $m_2 = 2.1 \pm 0.15$, $m_3 = 6.45 \pm 0.98$, $m_4 = 29.31 \pm 7.6$,
$C^2 = 1.07 \pm 0.02$, $P/M = 37 \cdot 10^3 \pm 109 \cdot 10^3$



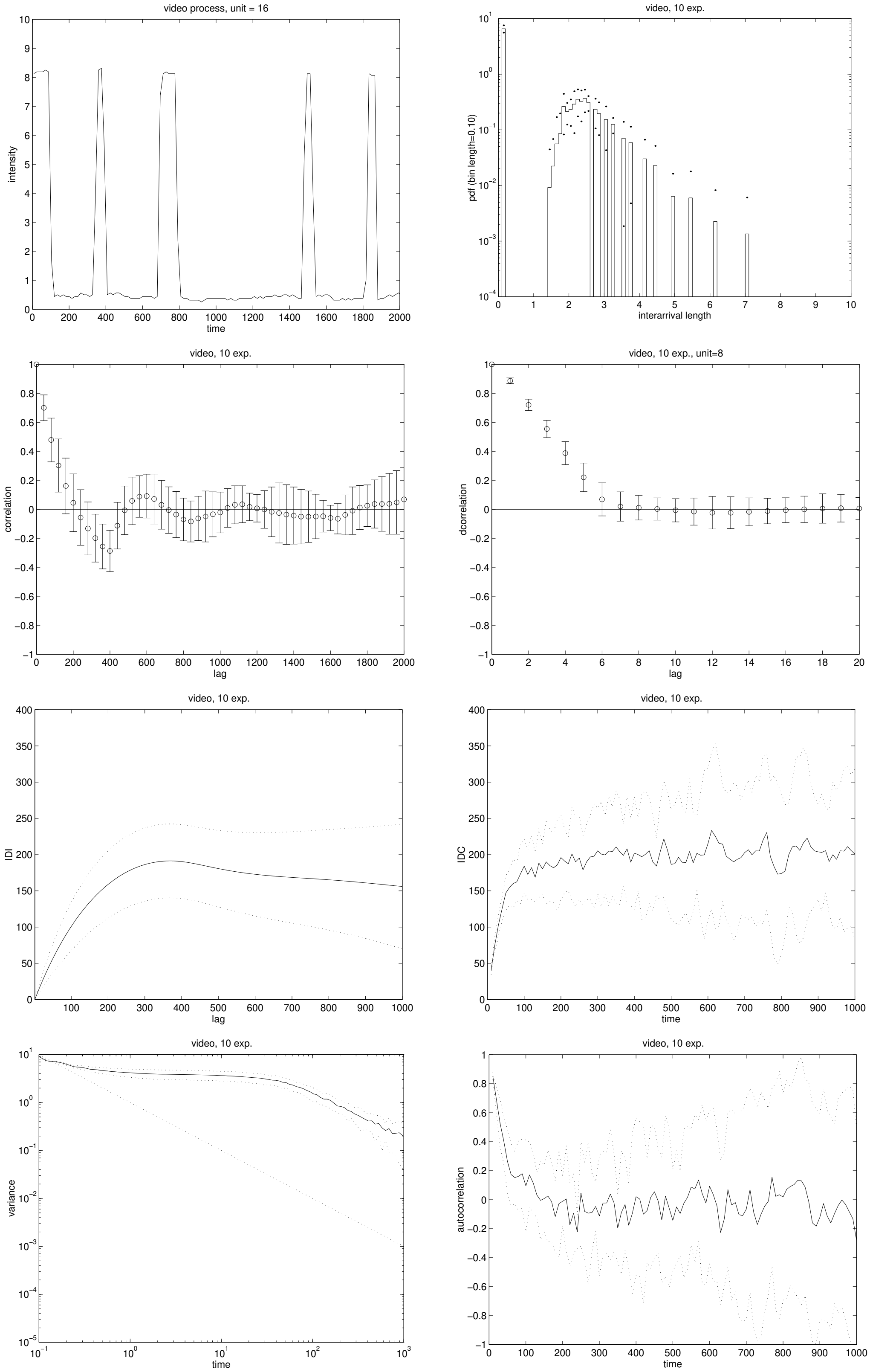
$\boxed{\texttt{voice}}$

$m_1 = 1.0000 \pm 0.0055$, $m_2 = 1.02 \pm 0.012$, $m_3 = 1.058 \pm 0.028$, $m_4 = 1.142 \pm 0.056$,
$C^2 = 0.0181 \pm 0.0043$, $P/M = 1.42 \pm 0.099$



$\boxed{\texttt{video}}$

$m_1 = 1.05 \pm 0.27$, $m_2 = 2.22 \pm 0.39$, $m_3 = 6.91 \pm 1.9$, $m_4 = 19 \pm 5.6$,
$C^2 = 1.42 \pm 0.48$, $P/M = 8.6 \pm 2.2$



B Generalized peakedness of traffic sources

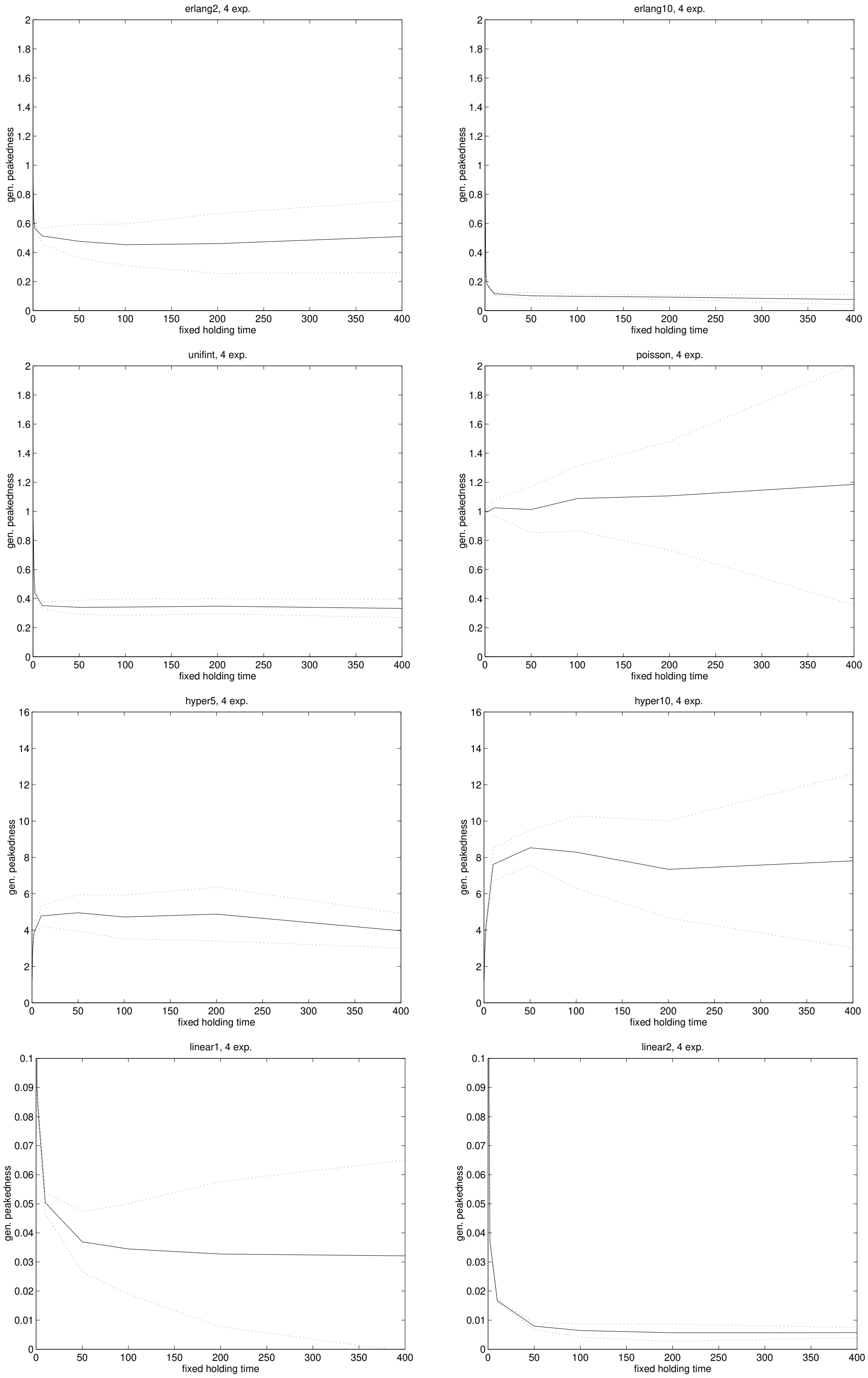
In the following pages I present the generalized peakedness of the 15 sample traffic sources as a function of the mean holding time. Four different holding time distributions are considered:

- deterministic (fixed holding time),
- exponential,
- hyper-exponential (with squared coefficient of variation 5),
- erlang-2.

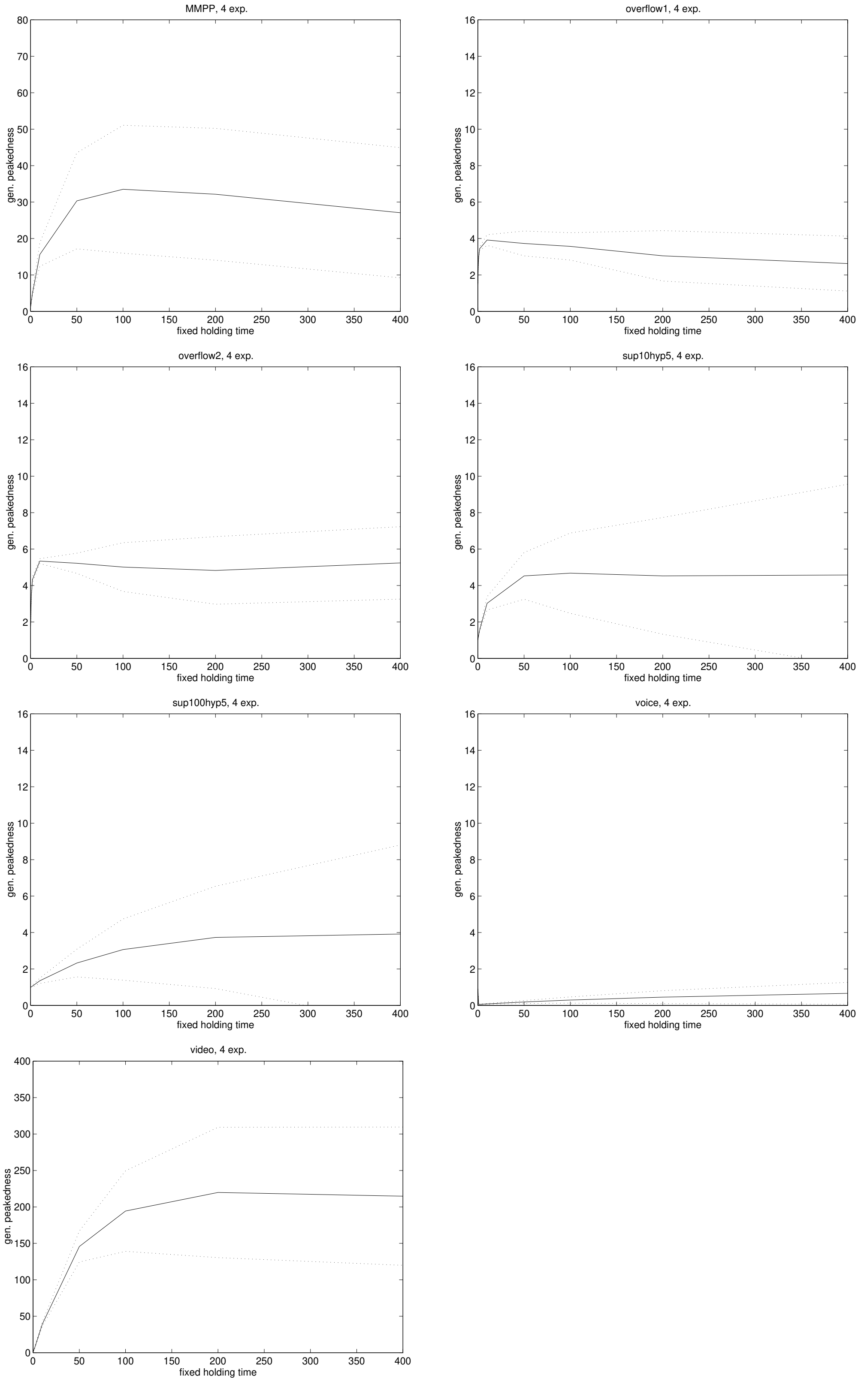
The graphs are shown in this order, two pages are devoted to each of the four holding time types. The graphs are based on simulations; each simulation involved 10000 interarrivals and each simulation was run four times. The 90% confidence intervals are shown on the graphs, but only as an indication of precision; more simulations are needed to establish better confidence intervals.

The graphs start from 1 and tend to a fixed value as indicated by the formulas of Subsection 17.4. Further research is needed to analyze the shape of the graphs.

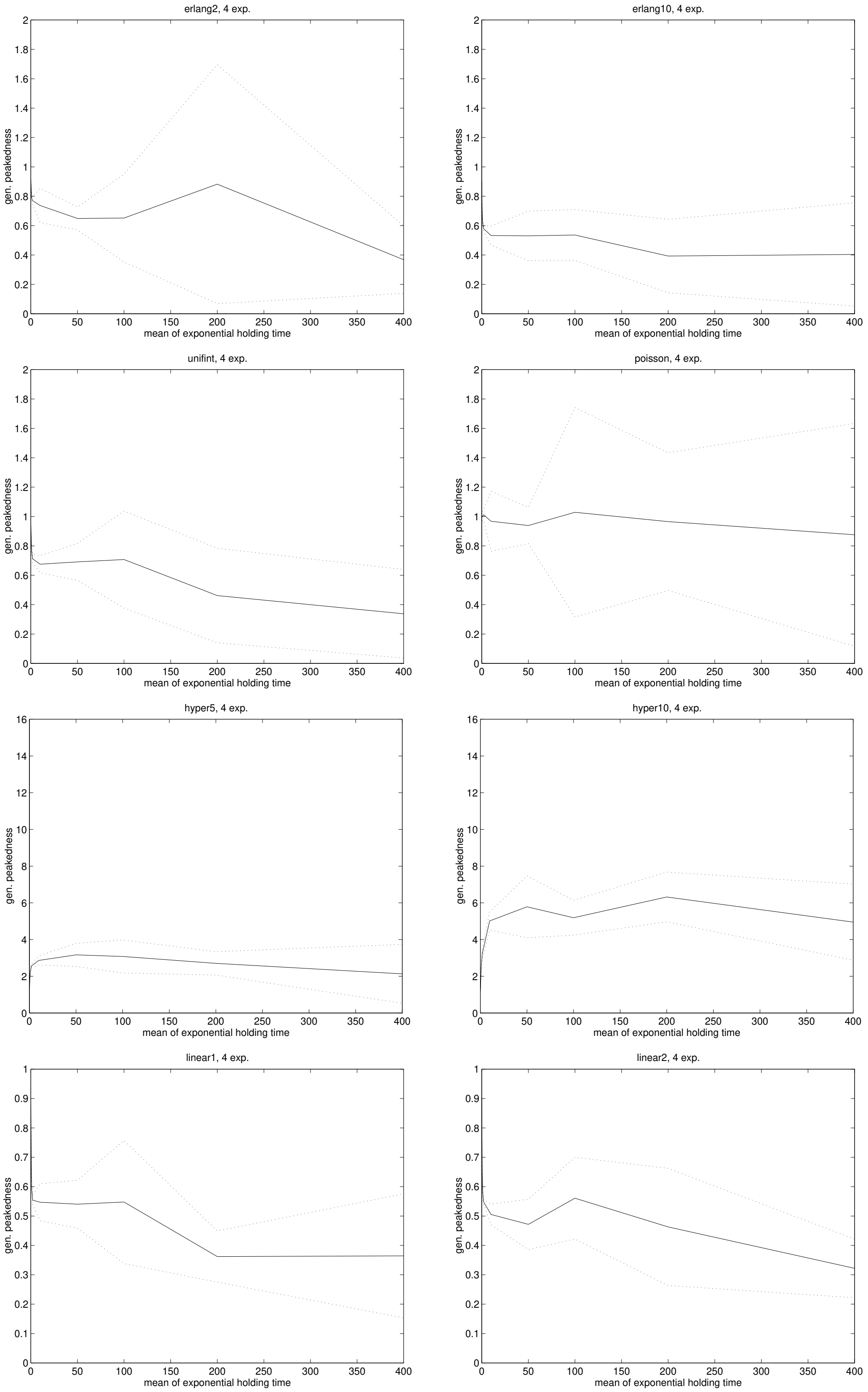
Deterministic holding time/1



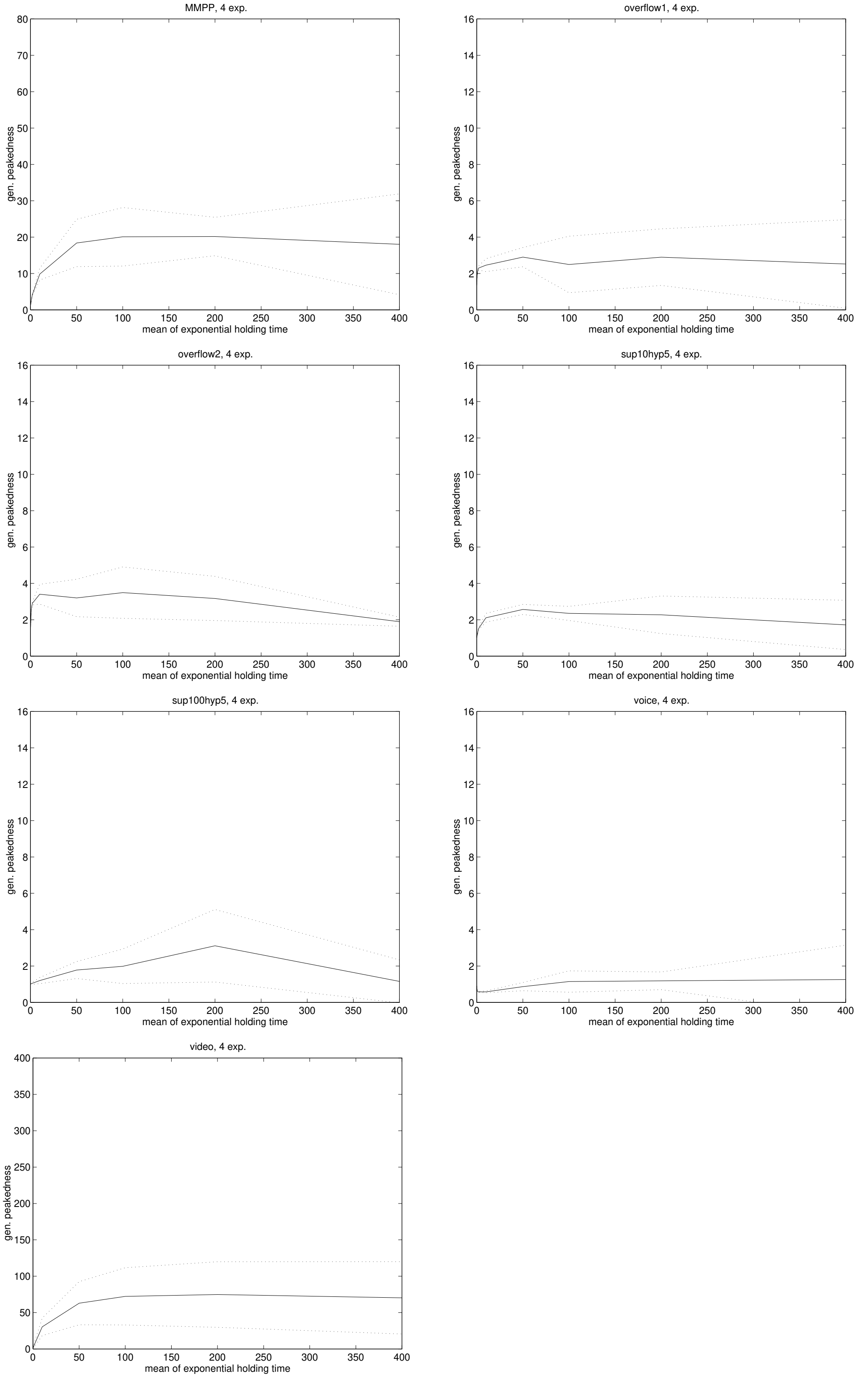
Deterministic holding time/2



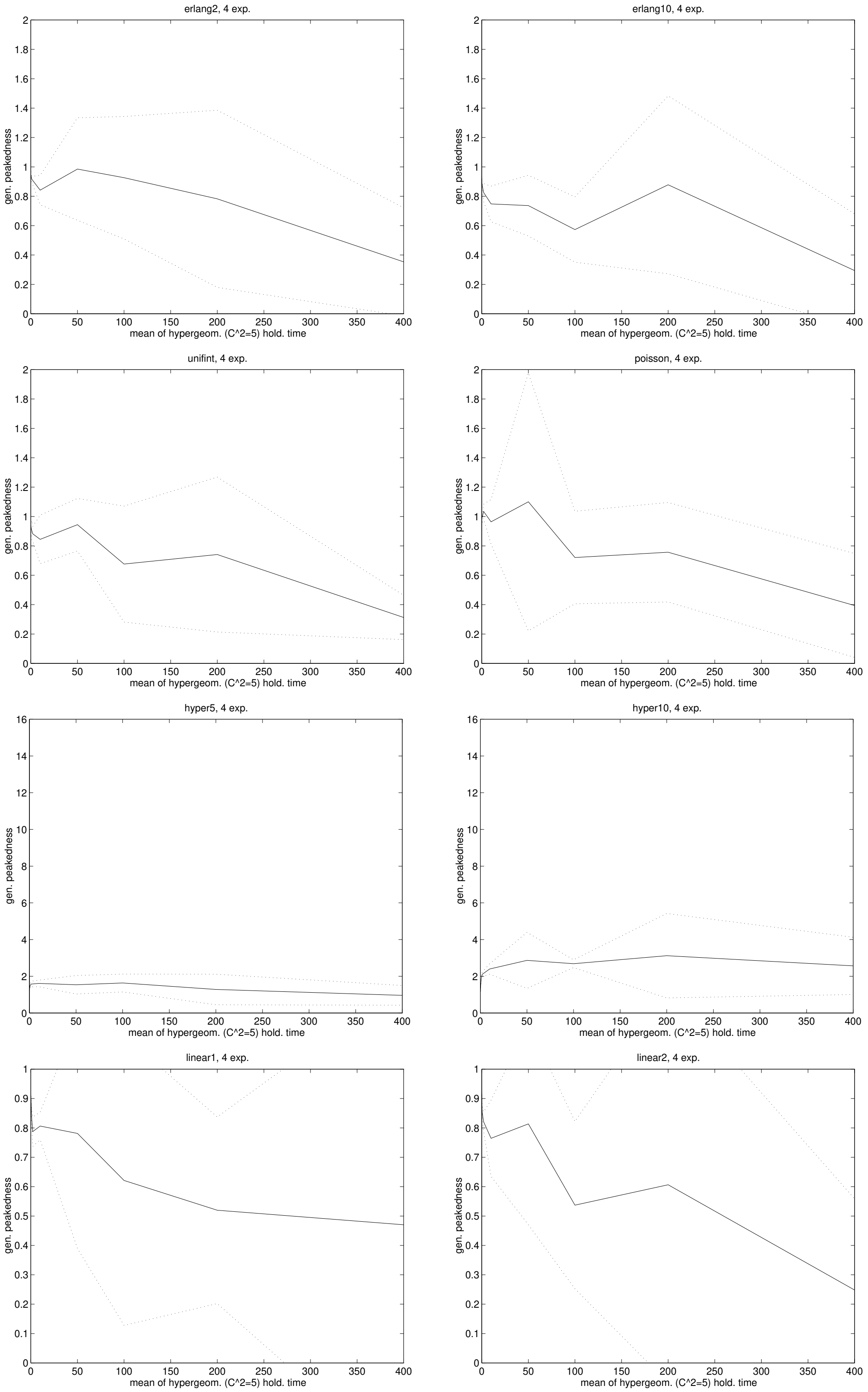
Exponential holding time/1



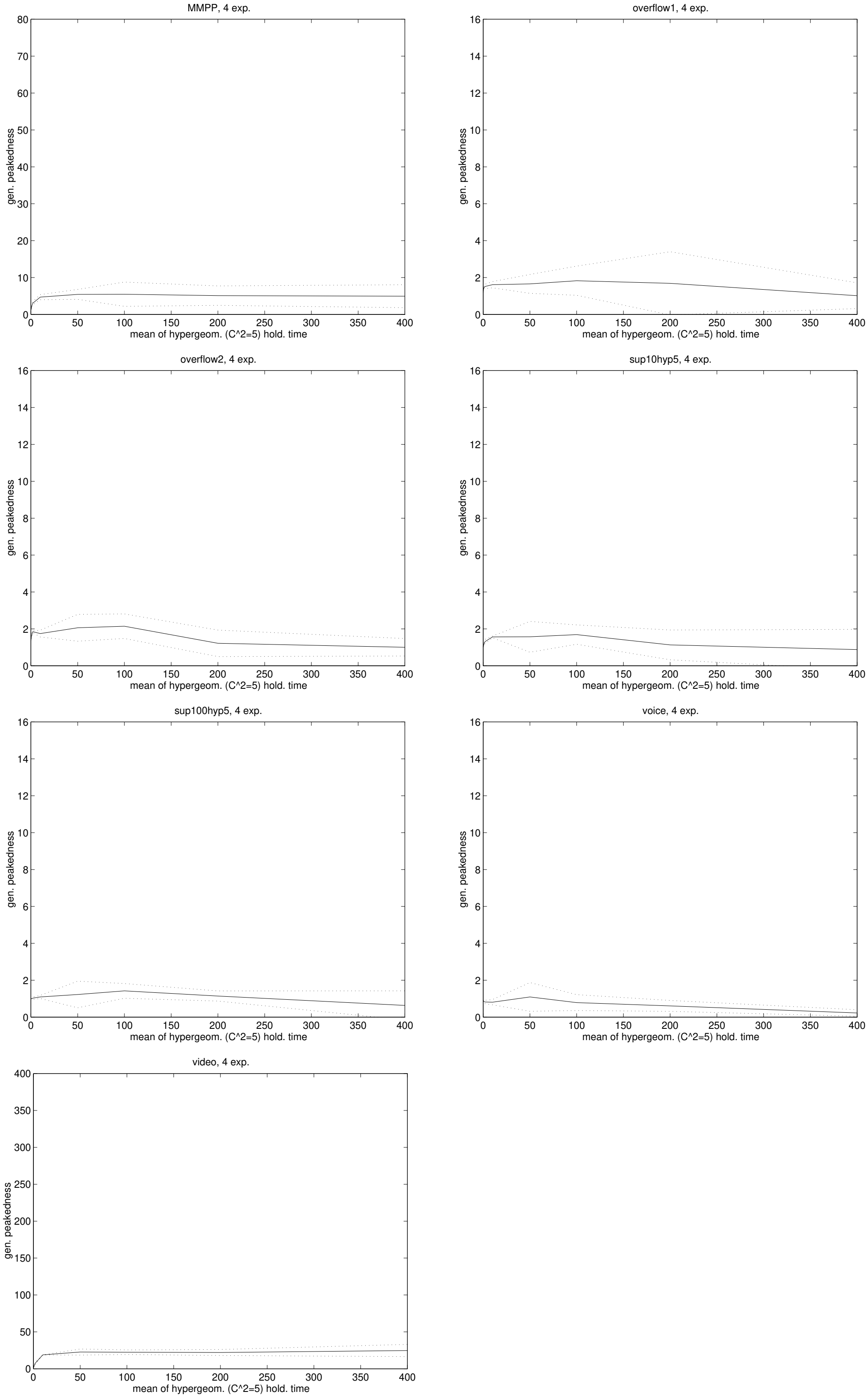
Exponential holding time/2



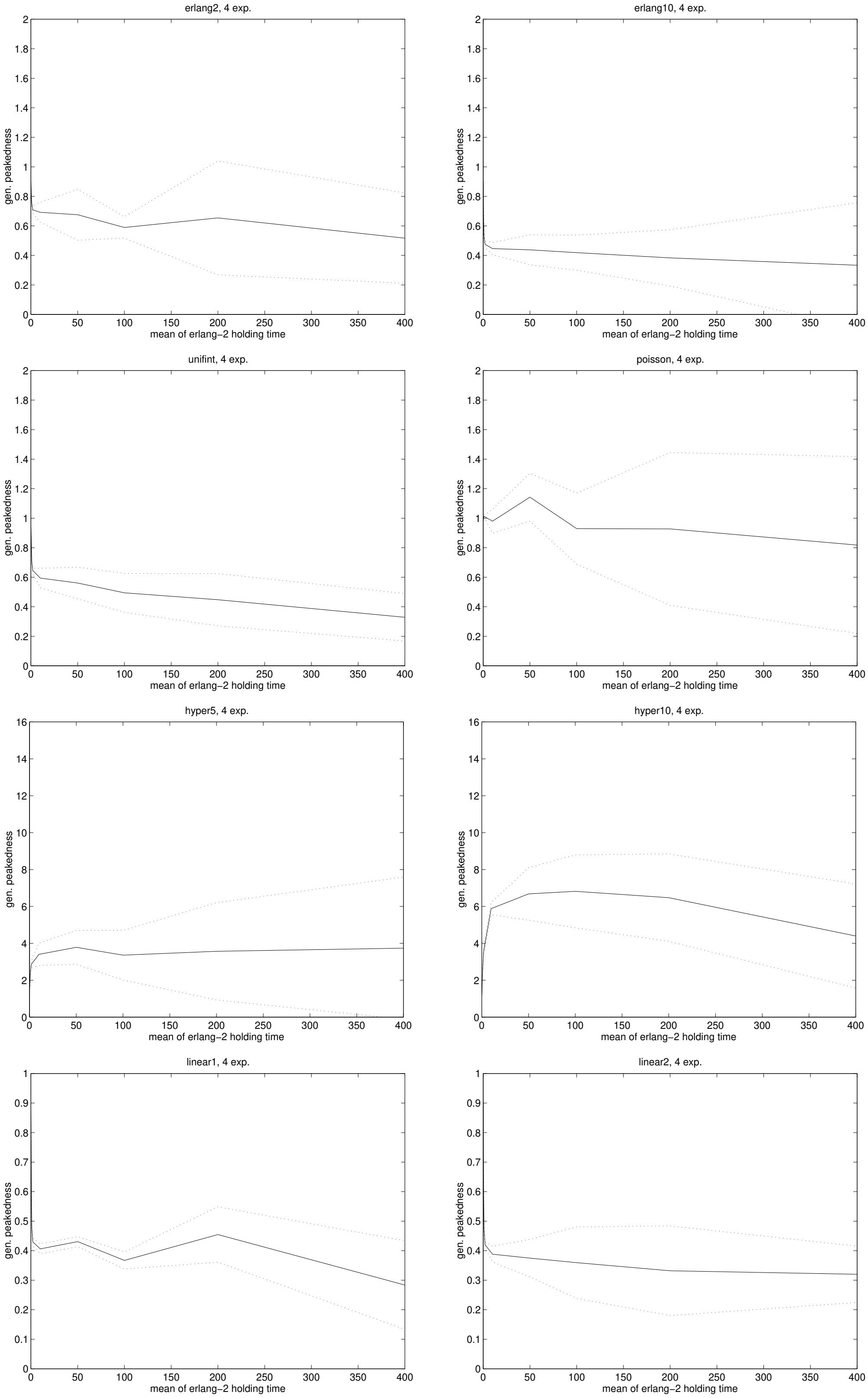
Hyper-exponential holding time/1



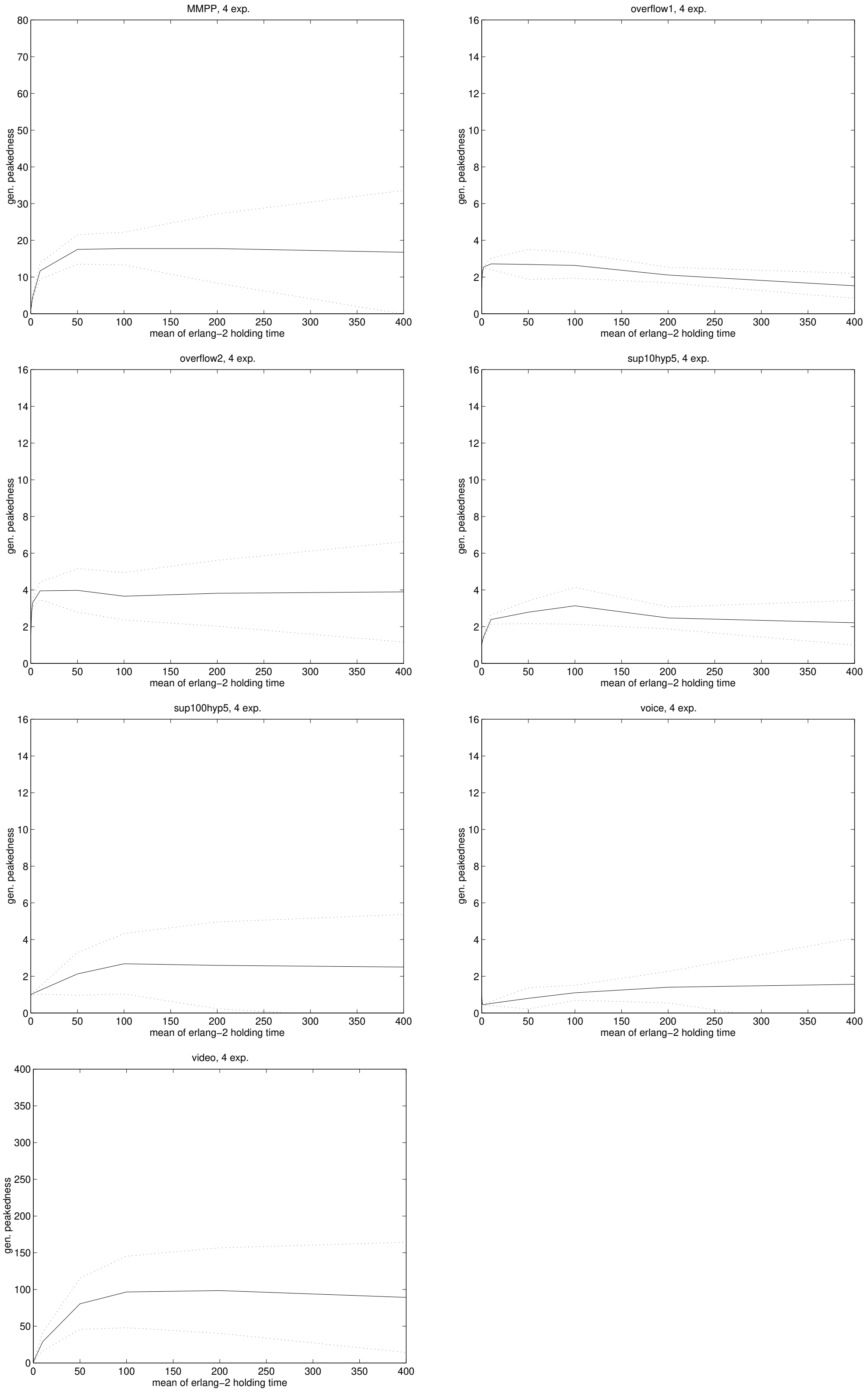
Hyper-exponential holding time/2



Erlang-2 holding time/1



Erlang-2 holding time/2



C Summary of programs

A summary of MATLAB routines written for this theses follows. A short description is given for each program. More help is available by typing HELP followed by the name of the program.

- Traffic generation and manipulation
  - General traffic generation:

`TRAFFGEN`	Traffic generator to produce sample traffic
`BATCH`	Generate batch process
`IPP`	Interrupted Poisson Process
`MMPP`	Markov Modulated Poisson Process
`TRAFF`	Example traffic to generate test results.

  - Practical traffic models:

`VOICEmodel`	Voice traffic model
`ARIPPmodel`	AR(1)+IPP traffic model for video traffic

  - Queue models:

`OVRFLOWQUEUE`	The overflow traffic of a system of type G/G/1/l.
`QUEUE0`	The number of requests in a system of type G/G/1.
`QUEUE`	Multipurpose FCFS queue modeling

  - Interarrival - counting process conversion:

`DTRAFF`	Discretize traffic

  - Traffic manipulation:

`COMB`	Comb several processes of interarrival time together.
`COUNTTAU`	Discretize time axis
`STARTPOINT`	Cut the first old elements in a vector.
`STARTAT`	Start process at a given time

- Traffic characterization
  - General traffic characteristics:

`SCV`	Squared coefficient of variation.
`CORR`	Correlation in discrete time
`CORREST`	Correlation estimate of a sequence.
`COVEST`	Covariance estimate of a sequence.
`IDC`	Index of Dispersion for Counts
`IDI`	Index of Dispersion for Intervals.
`VARTIME`	Variance-time curve estimation
`VARTIME1`	Variance-time curve estimation: method 1
`VARTIME2`	Variance-time curve estimation: method 2
`VARTIME3`	Variance-time curve estimation: method 3
`VTPLOT`	Variance-time curve divided by mt^2.
`P2M`	Peak to mean ratio
`GPEAKED`	Generalized Peakedness of a source.
`LINCOEFF`	Linear coefficients of a linear process computed

- Separating burst and inter-burst periods:

C2BURST	Cell level to burst level conversion.
R	Mean traffic intensity in burst-interburst couples

- Leaky buckets:

LEAKY	The number of cells in a leaky bucket
LEAKYEST	Estimate how a leaky bucket would react to the traffic
LEAKYGR	Leaky bucket graph for burstiness characterization
LEAKYRES	Show results of leaky bucket characterization

- Queue waiting time distribution:

WTIME_GP	Wait time distr. of C2/D/1 queue based on gen. peakedness
WTIME_SM	Wait time distr. of C2/D/1 queue based on simulation

• Displaying data

- Measuring and displaying:

Fig_src	Create and plot sample process.
MEASURE	Measure a characteristic for all the processes
DISPMEAS	Display measured data (mean and standard deviation)
MEASCORE	Needed by MEASURE.
GRAPH	Graph measure of a source
ALLGRAPH	Plot graph of a char. for all sources and save them.

- Name conversions:

SOURCE	Name of the nth source
FIGNAME	Get file name where a figure is saved
WHICHSRC	Get code of the source given by the name.

• General purpose

- Random number generation:

BINOM	Binomially distributed random variable
BINOMEST	Binomially dist. random var., estimated by norm distr.
FASTBINOM	Binomially distr. random variable: fast (stochastic routine)
RANDEXP	Returns random variables of exponential distribution.
GEOM	Geometrically distributed random number
ERLANG	Erlang distribution
RANDERL	Returns random variables of Erlang distribution.
HYPER	Hyperexponential distribution
COX2	Coxian-2 distributed random number
SGGEOM	Shifted Generalized Geometric distributed rand number.

- Combinatorics:

FACT	factorial
UNDER	n under k

- Statistics:

MOMENT	nth moment of data sequence.
WMEAN	Weighted mean
WSTD	Weighted standard deviation
LINREG	Linear regression

- Matrix manipulation:

`CUMULATE`	Cumulate all the elements in the left right corner.
`DIFFSEQ`	Differential sequence of a column vector.
`INFCONV`	'Infinite' convolution of two sequences.
`SHUFFLE`	Shuffle matrix: mix the elements.

- Plot:

`RIGHTSPREADPLOT`	Plot with horizontal steps spread to the right.

References

[1] György Miklós, Module Project Laboratory Report, Department of Telecommunications and Telematics, Technical University of Budapest, December 1995. *(In Hungarian)*

[2] György Miklós, "ATM Traffic Modeling", *Project Report*, Department of Communication Systems, Lund Institute of Technology, Lund, June 1996, CODEN: LUTEDX(TETS-7159) /1-32/(1996) & local 29.

[3] György Miklós, "ATM Traffic Analysis and Modeling", TDK '96 Student Conference, Technical University of Budapest, Faculty of Electrical Engineering and Informatics, Department of Telecommunications and Telematics, High Speed Networks Laboratory, October 1996.

[4] György Miklós, "Burstiness analysis of ATM traffic models", Conference of Graduating Students, Technical University of Budapest, Faculty of Electrical Engineering and Informatics, 1997.

[5] Sándor Molnár, György Miklós, "On Burst and Correlation Structure of Teletraffic Models", *5th IFIP Workshop on Performance Modelling and Evalution of ATM Networks*, West Yorkshire, U.K., 21-23 July, 1997.

[6] S. Molnár, I. Cselényi, N. Björkman, "ATM Traffic Characterization and Modeling Based on the Leaky Bucket Algorithm", *IEEE Singapore International Conference on Communication Systems*, Singapore, November 25-29, 1996.

[7] S. Molnár, I. Cselényi, N. Björkman, C-G. Perutz, M. Boda, "ATM Traffic Measurements and Analysis on a Real Testbed", *10th ITC Specialist Seminar*, Lund, Sweden, September 17-19, 1996.

[8] D. R. Cox, P. A. W. Lewis, "The Statistical Analysis of Series of Events", Methuen & Co Ltd, 1966.

[9] Vern Paxson, Sally Floyd, "Wide Area Traffic: The Failure of Poisson Modeling", *IEEE/ACM Transactions on Networking*, Vol. 3., No. 3., June, 1995.

[10] Will E. Leland, Murad S. Taqqu, Walter Willinger, Daniel Wilson, "On the Self-Similar Nature of Ethernet Traffic (Extended Version)", *IEEE/ACM Transactions on Networking*, Vol. 2., No. 1., February 1994.

[11] Faouzi Kamoun, M. Mehmet Ali, "Statistical Analysis of the Traffic Generated by the Superposition of N Independent Interrupted Poisson Processes" *Third Canadian Workshop, Information Theory and Applications,* Rockland, Canada, May, 1993.

[12] Victor S. Frost, Benjamin Melamed, "Traffic Modeling For Telecommunications Networks", *IEEE Communications Magazine*, March, 1994.

[13] A. E. Eckberg, "Generalized Peakedness of Teletraffic Processes", *ITC-10*, 1984.

[14] A. E. Eckberg, "Approximations for Bursty (and Smoothed) Arrival Queueing Delays Based On Generalized Peakedness, *ITC-11.*

[15] Sándor Molnár, "Evaluation of Quality of Service and Network Performance in ATM Networks", *PhD dissertation*, Technical University of Budapest, Department of Telecommunications and Telematics, 1995.

[16] G.L. Choudhury, D.M. Lucantoni, W. Whitt, "Squeezing the Most Out of ATM", *IEEE Transactions on Communications*, Vol. 44, No. 2, February 1996, pp. 203-217.

[17] Henk C. Tijms, "Stochastic Models: An Algorithmic Approach", John Wiley & Sons, 1994.

[18] Rainer Händel, Manfred N. Huber, Stefan Schrder, "ATM Networks: Concepts, Protocols, Applications", Second Edition, Addison-Wesley, 1994.

[19] Wolfgang Fischer, Kathleen Meier-Hellstern, "The Markov-modulated Poisson process (MMPP) cookbook", *Performance Evaluation 18* (1992) 149-171, North-Holland.

[20] D. R. Cox, W. L. Smith, "The superposition of several strictly periodic sequences of events", *Biometrika*, 40, 1-11.

[21] G. D. Stamoulis, M. E. Anagnostou, A. D. Georgantas, "Traffic source models for ATM networks: a survey", *Computer Communications*, vol. 17., number 6, June, 1994.

[22] J. Enríquez, A. López, J. Gamo, "Characterizing Alarm Traffic from Telephone Exchanges: Methodology and Case Study", ITC 14., Elsevier Science B.V.

[23] Hiroshi Yamada, Shuichi Sumita, "A Traffic Measurement Method and its Application for Cell Loss Probability Estimation in ATM Networks", *IEEE Journal on Selected Areas in Communications"*, Vol. 9., No. 3., April, 1991.

[24] Nasser M. Marafih, Raymond L. Pickholtz, Ya-Qin Zhang, "New Approximation For Analyzing the Multiplexing of Bursty Sources in ATM Networks", *Globecom '94*, San Francisco, USA, 28 Nov - 2 Dec, 1994

[25] R. Gusella, "Characterizing the Variability of Arrival Processes with Indexes of Dispersion", *IEEE Journal on Selected Areas in Communications*, February 1991, Vol 9., No. 2.

[26] Kazuyuki Onda, Kenji Nakagawa, "Approximation of Video Cell Traffic by AR(1) + IPP Model", *Electronics and Communications in Japan,* Part 1, Vol. 78, No. 8., 1995.

[27] Nasser M. Marafih, Ya-Qin Zhang, Raymond L. Pickholtz, "Modeling and Queuing Analysis of Variable-Bit-Rate Coded Video Sources in ATM Networks", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 4., No. 2, April 1994.

[28] Daniel P. Heyman, T.V. Lakshman, "Source Models for VBR Broadcast-Video Traffic", *IEEE/ACM Transactions On Networking*, Vol. 4, No. 1., February 1996.

[29] J. G. Gruber, 'A Comparison of Measured and Calculated Speech Temporal Parameters Relevant to Speech Activity Detection', *IEEE Trans. Communs.*, Vol. COM-30, No. 4, pp 728-738, 1982

[30] John Murphy, PhD report, 'Resource Allocation in ATM Networks', Dublin City University, March 1996,
WWW: http://www.eeng.dcu.ie/ murphyj/the/the/the.html

[31] Ulla Holst, Georg Lindgren, Jan Holst, Mikael Thuvesholmen, "Recursive Estimation In Switching Autoregressions with a Markov Regime", *Journal of Time Series Analysis*, Vol 15, No. 5, September 1994.

[32] David Barnes, Neil Smith, "An analysis of World-Wide Web Proxy Cache performance and its application to the modeling and simulation of network traffic.", *4th International Conference on Telecommunications Systems*, Modeling and Analysis, March 21-24, 1996, Nashwille, Tennessee, USA.

[33] Shuang Deng, Alan R. Bugos, "Access Traffic Model Inferred from WWW Servers", *4th International Conference on Telecommunications Systems*, Modeling and Analysis, March 21-24, 1996, Nashwille, Tennessee, USA.

[34] JS Kunicki and MF Scheffer, "Modeling Techniques For Packetized Traffic", *Proceedings of "III Polskie Sympozjum Ruchu Telekomunikacyjnego"*, Kraków, May 30-31, 1996.

[35] Josep Solé-Pareta, Jordi Domingo-Pascual, "Burstiness Characterization of ATM Cell Streams", *Computer Networks and ISDN Systems* 26, (1994) 1351-1363.

[36] Rene L. Cruz, "Service Burstiness and Dynamic Burstiness Measures: A Framework", *Journal of High Speed Networks* 2 (1992).

[37] Jaime Jungok Bae, Tatsuya Suda, "Survey of Traffic Control Schemes and Protocols in ATM Networks", *Proceedings of the IEEE*, Vol. 79., No. 2. February 1991.